International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 4 Issue 12, December-2015

Optimal Prescription Detection using C-ADR based on Contrast Set Learning with Expectation Maximation Algorithms

Roopa. M,
Research Scholar,
Department of Computer Science Karpagam Academy of Higher Education,
Coimbatore,Tamil Nadu,India

Dr. S. Manju Priya
Asst Proffesor
Department of Computer Science,
Karpagam Academy of Higher Education,
Coimbatore,Tamil Nadu,India

***Abstract*---Current healthcare industry has tremendous features and growth; due to this nature several medicines and its reaction finding processes need extra care and analysis. This type of drug reaction and prescription analysis is needed to equip with high absorption. The discovery of abnormal side effects that occur in more than one in a hundred and thousand patients are liable to signaled resourcefully by current drug surveillance techniques, but these same methods may take several time to generate the signals for abnormal side effects. In order to provide best and worst prescription finding from huge dataset, the system proposes an innovative data mining framework. This work initially performs contrast set learning algorithm for drug and symptom frequency detection. Rules from contrast set learning algorithm will be applied with multi objective algorithms such as casual leverage using EM (Expectation Maximization) algorithm. This helps to mine the best and worst drug pairs and symptoms from huge patient electronic treatment dataset.**

Keyword---Adverse Drug Reaction (ADR); causal-leverage; C-ADR (chronological adverse drug reaction); DDR (drug-to-drug relations); expectation maximation (EM)

1. INTRODUCTION

Modern healthcare systems generates large amount of information, due to its huge size, It is necessary to extract useful knowledge and providing scientific decision making for the diagnosis and treatment of disease from the database increasingly becomes necessary so several data mining techniques have been proposed.[1]It can also improve the health care domain and promote the development of healthcare domain. Adverse Drug Reaction (ADR) is the unexpected or abnormal effect resulting from the use of a drug or the use of two or more drugs. The occurrence of an ADR is called an Adverse Event (AE). Adverse drug reactions (ADRs) symbolize a severe disaster wide-reaching. The term "adverse effect" is preferable to other term such as side effect. They can complicate a patient's medical condition or contribute to increased morbidity, even death. Adverse Drug Reaction often refers to the side effects caused by a drug and can be categorized based on the mechanism by which they are caused. Their knowledge and understanding is necessary for practitioners to monitor drug therapy and ADR detection.

Survey have shown that ADRs contribute to about 5% of all hospital admissions and represent the fifth most common cause of death in hospitals. Interestingly, it was found that more than 70% of these ADRs were potentially avoidable. A more recent study in Brazil suggests ADRs may be the cause of an even higher proportion of hospital admissions for the elderly as it showed that ADRs were the cause of hospitalization for more than 50% of elderly patients. [2]

2. PROBLEM DEFINITION

The ADR and negative prescription finding generates a huge burden for the healthcare service in terms of causing both patient morbidity and mortality and costing large sums of money. So finding best, worst prescriptions and the evaluation of drugs with its normal and abnormal symptoms is necessary. Discovering this kind of causal relationships can help us prevent or correct negative outcomes caused by its antecedents.[3] However, mining these relationships is challenging due to the difficulty of capturing causality among trials and the infrequent nature of the events of interest in these applications. Finding causal reactions and relations between two medicine prescriptions or sets of events with relatively low frequency is very useful for various processes in different domains. This should be done with high accuracy, because finding negative and abnormal side effects caused by prescribed medication is very important. In this paper, the system employs a knowledge based approach to capture the degree of causality of an event pair within each sequence since the resolve of causality is habitually in due course of application or domain dependent. The study was motivated by the need of discovering ADR and casual signals in post-marketing examination of drugs.

3. METHODOLOGY

A. Existing system:

There are several approaches implemented in existing to detect the adverse of drug reactions.

- Fuzzy based system has been implemented in existing. That failed to bring the accuracy.
- Several top down and bottom up approaches proposed to detect of rare association rules are based on traditional interestingness measures like support and confidence. [4]

- One consequence of these actions is that they merely discover the geometric correlation between two points.
- They do not point out any chronological relationship between X and Y. In addition, they are not capable to confine the contributory relationships between two event sets.[2]

B.Drawbacks

Existing system may be challenged to convey their acquaintance in the form of probability distributions.

- Failed to compare two diverse blend treatments along with the temporal incidence of drugs and drug demanding reactions and changes in order to forecast the most valuable treatment.
- Proof of each drug reaction based on various circumstances is very difficult.
- The existing system provides only an estimated solution for treatment selection, so that is not an accurate system.
    - Prior Knowledge need
    - Slow in reaction finding

C.Proposed System

Developing and incorporating an exclusion mechanism that can effectively find the undesirable prescriptions caused by frequent events. The proposal aims at introducing an exclusive chronological causal-leverage measure for sequential drug reaction analysis. The system aims to propose a data mining algorithm to mine C-ADR (chronological adverse drug reaction) from electronic patient database based on the new measure. The system proposed a new method for analyzing and comparing the drugs for best and worst drug detection. This also helps to identify the best and worst prescriptions and drug pairs. The main objective of the proposed technique is as follows.

- Best finding of drugs, based on the drug reactions with temporal reactions.
- This aims at developing an accurate analysis of best drug finding.
- With the use of causal-leverage with EM the proposed system improves the accuracy.
- It aims to provide a constrained optimized drug selection.

Finding causal and rare associations between two measures or sets of measures with relatively low rate of recurrence is very useful for various real-world applications such as medical domain. A drug used at a suitable dose may origin one or more adverse drug reactions (ADRs), even though the probability is less. Providing this kind of causal relationships and suggestion for best drug finding can help the user to avoid or correct negative outcomes caused by its antecedents. Mining the adverse drug relationships and DDR (drug-to-drug relations) are tough due to the difficulty of capturing causality among events. ADR is also hard to implement due to the insufficient training datasets in the medical domain.

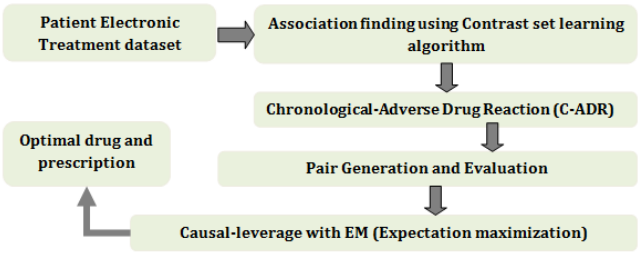

Fig1: Overall structure of the proposed work.

The process of data mining in prescription analysis is performed in the following three steps:

Step 1: Reduce the complexity of the algorithm. To this purpose, the drug names are mapped to their corresponding codes. This will also assist in reducing redundant verification and will strengthen the result finding.

Step 2: A set of candidate drug-drug adverse drug events associations was generated using the association rule mining.

Step 3: These associations are filtered to remove false associations to identify potential features.

i. Association finding: (Contrast Set Learning Algorithm)

Contrast set learning is a form of association rule learning, which seeks to identify meaningful differences between separate drug groups by reverse-engineering where the key predictors can identify for each particular group. This finds the different support valued dataset form the database. For example if "drugA" has elevated confidence than additional drugs, this consequences as "drugA" is a contrast set cluster medicine. The contrast set learning algorithm helps to discover that the major difference between different drugs related details. The association finding segment helps to trail all the support and confidence of the drugs.

Initially this process performs the following cycles.

- Scrutinize the database
- Candidate generation
- Support and confidence calculation
- Contrast set grouping

The frequency detection is the process of identifying the total number of occurrences of prescriptions with various factors. For example, the drug prescriptions for every patient and causal and other symptoms will be captured at every time interval should be identified by the following formula (1).

$$\text{Support}(X) = (T(x))/n \qquad (1)$$

Where x is an attribute, T(x) is the total number of occurrences of x and n is the total transaction in the dataset. In the proposed system the prescriptions data's are composed and applied for frequency discovery.

Tid	Data's	Occurrences/total transaction
1	Fever, cold	Total support : 6 Support of(vomit)=(2/6)=33.3%
2	Vomit	
3	Vomit	
4	Headache	
5	Drowsy	
6	Drowsy	

Table1: Support calculation of symptom type

In order to identify the frequency different attributes are considered and applied into the formula (1), thus the ultimate yield will be useful to the association data. As like the support additionally, the confidence among different attributes is considered by the following formula (2).

Confidence (X➔Y) =P(Y/X) (2)

Where x is an attribute, (Y) is the total number of occurrences of Y and (X) is the total number of occurrences of X is the total transaction in the dataset. This work prescriptions data's are collected and applied for frequency detection. The rule generation from the confidence is important to analyze the association between attributes. X➔ Y holds with confidence c if c% of the dealings in D that contain X also contain Y. Rules greater than a user specified confidence is calculated to have minimum confidence from the dataset.

Tid	Attributes(drug, symptom, other attributes)	Given X➔ Y Confidence=Occurrences of (Y)/ Occurrences of (X)

Table 2: Confidence calculation of drug type

Items are pruned from the tree when all specializations of the node can never lead to a major and bulky contrast set. The outcome is order based on:

- The minimum divergence size: The maximum difference between the supports of any two groups but it is greater than a user specified threshold.
- Expected drug frequencies: The expected drug frequencies of a transaction table can only decrease as the contrast set is focused. When these frequencies are too little, the validity of the test is debased.
- Limits: An upper bound is kept on the distribution of a statistic calculated when the null hypothesis is true. Items are shortening when it is no longer possible to meet this cutoff.

This work treats the problem of mining contrast-sets as a tree search problem in drug analysis. The root node is an empty contrast set, and we generate leaf of a node by utilizing the set by adding one more term. We use a canonical ordering of attributes to avoid visiting the same node twice. Leaves are formed by appending terms that follow all existing terms in a given ordering. This finds the different support valued dataset form the database. For example if "drugA" has high confidence than other drugs, this results as "drugA" is a contrast set group medicine.

Contrast Set Algorithm (CSA) is projected to stumble on all contrast sets whose support differs significantly and transversely on collections. A contrast set is represented as an item set, where a group or collection is the given dataset D where contrast sets occur. Groups are represented by items

E.g. Hi, Hj, typically attribute values. Formally, the goal is to find those contrast sets (CSA) where:

$$P(csa|Hi) \neq P(csa|Hj) \quad (1)$$

and $$\max_{i,j} |\sup(csa,Hi) - \sup(csa, Hj)| \geq T \quad (2)$$

Where T is threshold and called the minimum support. Support in this scenery deals the relative prevalence of the contrast set within a group (Hi and Hj). Contrast sets where equation 1 is statistically valid are called important and contrast sets where equation2 is met is referred as bulky. If mutually necessities are achieved, then the contrast set is measures a digression. The statistical significance standard ensures that the contrast set represents a true divergence between the medicines. The second standard measures the effect size and ensures that the whole thing that is reported to the user provides a large enough effect to be important.

ii. C-ADR

Chronological Adverse Drug Reaction (C-ADR) is the unexpected or unwanted effect resulting from the use of a therapeutic drug or the use of two or more drugs. The occurrence of an ADR is called an Adverse Event (AE). The chronological adverse drug reaction finding is the improvement of ADR, which facilitates the fast analysis of the Adverse Event data with time segmentation. C-ADR calculates the extent of causality between symptoms and some of its adverse effects which are used by the real patients to form simulated patient cases and also containing drug-symptom pairs of interest with assorted degrees of causality. The model's legality was then established by comparing the decisions prepared by the model and those by two self-determining skilled physicians for the set of simulated patients.

C-ADR Algorithm

Initially the system proposed a data mining algorithm to mine C-ADR signal pairs from electronic patient database based on the new and predictive measure. The C-ADR helps to identify the side effects of a particular drug. Based on that the system, it declares the best and the worst drug pairs from the dataset.

Drugs and their allied C-ADRs have causal interaction. In this sector, how to examine search for potential C-ADR signal pairs from an electronic patient database using the above chronological causal-leverage measure is elaborated. This assumes that patient data are stored in relational tables in a database and can be retrieved using database language like structured Query Language (SQL).

Steps

Step 1: A list of all drugs D (d1; d2; . . . ; dm) in the database.

Step 2: Calculate and return the support count for each drug.

1. S=Support(d1)=(no of dq in dataset D/total no of dataset D)* 100
2. Return S.
3. List of all symptoms S (s1; s2; . . . ; sn) in the database and the support count for each symptom using the step 2.
4. The lists of drugs and symptoms are needed to form all possible drug-symptom pairs whose causal strengths will be assessed.
5. Finds chronological adverse drug reaction pair.

Since a patient database normally contains only a subset of all drugs on the uphold and a subset of all symptoms, it is obligatory to investigate the Patient Drug Table and the Patient Symptom Table to acquire the drugs and symptoms enclosed by the database. The following figure shows the overall frequencies of C-ADR for treatment1.

treatement	frequency
fever	11
achy_muscles	10
fatigue	30
sore_throat	34
nausea	40
skin_rash	44
weight_loss	41
dry_cough	32
pnemonia	35
night_sweats	27
nail_changes	28
yeast_infections	24
confusion	19
cold_sores	14
tingling_weakness	13

Fig2: Frequency calculation of C-ADR for treatment 1

The above algorithm used to discover the Patient Symptom Table for a list of symptoms covered by the database as well as the support count for every symptom, which stated in the above table. If users are only interested in mining the prospective C-ADRs of a exacting drug or a pair of drugs, the users can denote the drugs of awareness. Likewise, the users can also denote the list of symptoms if they want to analyze which drugs can cause the symptoms of interest. In mutual cases, though 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.

	drug	symptoms	freq		drug	symptoms	freq
1	1	achy_muscles	3	21	11	night_sweats	1
2	1	cold_sores	3	22	12	nail_changes	1
3	1	confusion	6	23	2	achy_muscles	3
4	1	dry_cough	6	24	2	cold_sores	4
5	1	fatigue	2	25	2	confusion	3
6	1	fever	1	26	2	dry_cough	9
7	1	menstrual	3	27	2	fatigue	10
8	1	nail_changes	6	28	2	fever	5
9	1	nausea	7	29	2	menstrual	2
10	1	night_sweats	4	30	2	nail_changes	2
11	1	pnemonia	11	31	2	nausea	7
12	1	skin_rash	8	32	2	night_sweats	6
13	1	sore_throat	7	33	2	pnemonia	9
14	1	tingling_weakness	6	34	2	skin_rash	13
15	1	weight_loss	9	35	2	sore_throat	8
16	1	yeast_infections	5	36	2	tingling_weakness	1
17	10	nail_changes	1	37	2	weight_loss	9
18	10	nausea	1	38	2	yeast_infections	4
19	10	night_sweats	1	39	3	achy_muscles	2
20	11	nail_changes	1	40	3	cold_sores	4

Fig 3: Drug and its ADR with the frequency calculation

From the above figure, the drug and its symptoms has been collected with its frequencies. Using the above dataset the adverse drug reaction can be found without any training samples.

iii. Pair Generation and Evaluation

Every drug and its symptoms are evaluated together. This process extracts all item sets i.e. drugs and its combined reactions for evaluation. This step is the progression for pair generation and evaluation. In this algorithm drug-symptom pairs can be straightforwardly generated. The pairs are drug-drug pairs, drug-symptoms-time pairs, symptom-symptom pairs or combinations of various drugs and symptom pairs. So this algorithm generates a much fewer number of candidate rules, which imply much less complexity.

Example:

Lisinopril→cough.

But the current problem demands excavation of multiple associations like:

Aspirin + Warfarin → bleeding or

Azithromycin+ Fleroxacin → sleep disorder + abdominal pain

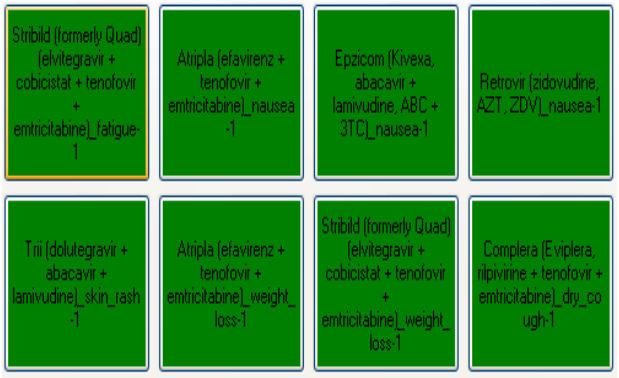

Fig 4: Pair generation for drugs with commonly occurring symptoms

These have been calculated with the help of contrast set learning algorithm. The PET database is huge and the rules to be explored corresponding to two or more drugs is computationally problematic. To improve the efficiency of the algorithm and reduce search space, additional criteria to increase support and confidence of the association rules and the convention with high precedence, such as rules containing certain number of items or set of items, are enforced in finding DDRs (Drug to Drug Reactions). The general contrast set learning algorithm which is optimized for the above criteria is implemented. This process also creates an additional pair drug-symptoms- period pair, which considers the ADR with time basis.

The Contrast Set algorithm uses the downward closure property of frequency to diminish the investigation space of association rules. In the process of DDR mining this means that if some combination of drugs and DRs (Drug Reactions) are occasional, then the superior set of combinations which gets formulated on the infrequent one will also be intermittent and can be eliminated from evidence.

iv. Causal-Leverage with EM (Expectation Maximization)

The EM method for discovering maximum likelihood or maximum a posteriori (MAP) estimates of parameters in statistical models. This has been combined with casual leverage. For individual symptom and drug pair the likelihood value will be calculated. This algorithm explains how to figure the causal leverage value of a general pair between event A and B. Both A and B could be either drug event or symptom event. Initially the drug or symptom hash table is searched in order to get the support count for event B. Then for each PID that supports the itemsets and a process called cue abstraction is used to extract a set cue values T from the related patient case. Specifically, a list of drug start dates and a list of symptom dates are restored from the Patient Drug Table and the Patient Symptom Table respectively. Finally rank all the pairs in a decreasing order according to their chronological causal-leverage values after all these values are computed. The expectation maximization (EM) algorithm enables parameter estimation in probabilistic models with incomplete data.

Consider a simple drug analysis experiment in which a pair of drugs A and B of unknown reaction pair, θA and θB, respectively (that is, on any given treatment, drug A will produce a symptom "vomit" with probability θA and "head ache" with probability 1–θA and similarly for drug B). The goal is to estimate θ = (θA,θB) by repeating the following procedure five times: randomly choose one of the two drugs (with equal probability), and perform ten independent drug combinations with the selected drug. Thus, the entire procedure involves a total of 50 drug experiments (Fig. 1a). During the experiment, deduce that we keep track of two vectors x = (x1, x2, …, x5) and z = (z1, z2,…, z5), where xi ∈ {0,1,…,10} is the number of C-ADR observed during the ith set of experiments, and zi ∈ {A,B} is the identity of the drug used during the ith set of drug. Parameter evaluation in this scenery is known as the complete data case in that the values of all relevant random variables in the proposed model (that is, the result of each drug reactions and the type of drug used for each experiment) are known. Here, a simple way to estimate P (A) and P (B) is to return the observed proportions of reactions for each drug:

$$P(A) = \frac{No\ of\ C-ADR\ using\ drug\ A}{Total\ no\ of\ ADR}$$

$$P(B) = \frac{No\ of\ C-ADR\ using\ drug\ B}{Total\ no\ of\ ADR}$$

In this perceptive maximum likelihood estimation method assesses the quality of a drug analysis model based on the probability it assigns to the analyzed data.

Optimal prescription finding process

The best drug detection using C-ADR with EM and CSA algorithms adopts a strategy consisting in selecting the appropriate ADR of the overall set of ADRs.

Input: D – Patient data

T_N – Treatment

P – Pair generation attributes

C – Drug Reactions DR

S_c – contrast set group

Step 1: Read dataset from D

a) Read the attributes and values from T_N

b) Every attribute is set into 'P"

c) Set of ADRs denoted as C

Step 2: Pre- process steps

Step 3: support and confidence calculation

a. identify base rules for every P from the following step

b. Sup (P) = (T(P))/n , Confidence (Pi➔Pi+1)

Step 4: find contrast set group Sc.

a) Set S_c.
b) If the property is already in the ADR- find the probability
c) Else if new ADR, perform the C-ADR and EM

Step 5: check the threshold and probability values

Step 6: detect the prescription ratio from the dataset and return best prescription and drug.

The system performs the detailed analysis based on the temporal analysis of the drugs. Finally the system performs the analysis and results of the proposed system. Then the system finds the Best Drug and worst drug which is associated with new symptoms using the above algorithms.

V. Result Summary

The result shows the proposed methods has successfully implemented with high accuracy and low computational overhead. It is known from the implementation and analysis, with 35 drugs, drug number 1 ie., Atripla (efavirenz + tenofovir + emtricitabine) is resulted as best drug and the worst drug is 12 which showed more ADR and this has been identified using the C-ADR and EM algorithms with several patient records.

4. CONCLUSION

ADR finding is a difficult task and to find out which drug causes abnormal side effects can be done only with surveillance techniques which are a prolonged process so to find best and worst drug, this system uses an EM and chronological adverse drug reaction approach to understand different drug reaction and changes with different ADR signal pairs. The proposed system also implements the expectation maximization algorithms for better and accurate prediction of drug and drug reactions. C-ADR calculates the extent of causality between symptoms and some of its adverse effects which results in accurate prediction of drug and drug reactions, thus the proposed system can give DDR and Drug-Symptoms for finding ADR.

5. ACKNOWLEDGEMENT

I thank the Karpagam University for the motivation and encouragement for giving the opportunity to do this research work as successful one.

6. REFERENCES

[1] Sandhya Joshi, Hanumanthachar Joshi, “Applications of data mining in health and pharmaceutical industry”, International Journal of Scientific & Engineering Research, Vol- 4, Issue- 4,pp- 915,2013.

[2] Jenna Reps, Jonathan M. Garibaldi, Uwe Aickelin, Daniele Soria, Jack E. Gibson, Richard B. Hubbard, A Novel Semi-Supervised Algorithm for Rare Prescription Side Effect Discovery, IEEE Journal Of Biomedical And Health Informatics, vol. 18, no. 2, pp-537-547,2014.

[3] Ji Y1, Ying H, Dews P, Mansour A, Tran J, Miller RE, Massanari RM., “A Method for Mining Infrequent Causal Associations and Its Application in Finding Adverse Drug Reaction Signal Pairs”, IEEE transactions on Knowledge and Data Engineering,, Volume:25 Issue:4,pg-721- 733,2012.

[4] S. Prakash and S. Kanjanadevi,“Post Market Drug Analysis using Irregular Pattern Mining Scheme”,International Journal of Innovative Research in Computer and Communication Engineering, Vol-2, Special Issue 1,pp-3838-3844, 2014.

[5] Yihui Liu and Uwe Aickelin,”Detect adverse drug reactions for the drug Pravastatin”.pg-1188-1192, 2012.

[6] Roseleen Vino.I and Kerana Hanirex.D “Sharing ADRs for Immediate Treatment”, (IOSR) Journal of Computer Engineering (IOSR-JCE), Vol-3, pg-62-65, 2011.

[7] R. Sindhulakshmi and, D. SaravanaPriya, “A Data Mining approach for efficient event analysis to find Adverse Drug Reactions”, International Journal of Research in Computer Applications and Robotics, IJRCAR, Vol.2 Issue.3, Pg.: 103-109,2014.

[8] S. Prakash and S. Kanjanadevi,“Post Market Drug Analysis using Irregular Pattern Mining Scheme”,International Journal of Innovative Research in Computer and Communication Engineering, Vol-2, Special Issue 1,pp-3838-3844, 2014.

[9] S.Sandhya Lakshmi and S.Sheba Roshni,“Mining Adverse Drug Reaction For Infrequent Causal Association”, International Journal of Computer Science and Engineering Communications- IJCSEC, Vol.2 Issue.4, pg-415-419, 2014.

[10] Vikask Kumar Saw and Kumaran U, “An Optimal Approach for Mining Rare Causal Associations to Detect ADR Signal Pairs”, International Journal of Scientific and Research Publications, Volume-4, Issue-5, 2014.

[11] Ji Y1, Ying H, Dews P, Mansour A, Tran J, Miller RE, Massanari RM., “A potential causal association mining algorithm for screening adverse drug reactions in postmarketing surveillance”, IEEE transactions on information technology in biomedicine, Vol-15, No-3,pg- 428-37,2011.

[12] Deforche K, Camacho R, Grossman Z, Silander T, Soares M, Moreau Y, et al. Bayesian network analysis of resistance pathways against HIV-1 protease inhibitors. Infection, Genetics and Evolution;Vol-7,Issue-3,pp-382–90, 2007.

[13] Deforche K, Silander T, Camacho R, Grossman Z, Soares M, Van Laethem K, et al. Analysis of HIV-1 pol sequences using Bayesian networks: implications for drug resistance. Bioinformatics;Vol-22,Issue-24,pp:2975–9, 2006.

[14] Beerenwinkel N, Rahnenführer J, Däumer M, Hoffmann D, Kaiser R, Selbig J, et al. Learning multiple evolutionary pathways from cross-sectional data. In: Bourne PE, editor. Proceedings of the eighth annual international conference on research in computational molecular biology. San Diego, CA, USA: ACM; p.p- 36–44, 2004..

[15] Pearl J. Probabilistic reasoning in intelligent systems: networks of plausible inference. San Francisco, CA, USA: Morgan Kaufmann;pp:1-28, 1988.

[16] 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.

[17] 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.

[18] Huidong Jin, Jie Chen, Chris Kelman, Hongxing He,Damien McAullay, and Christine M. O’Keefe, “Mining Unexpected Associations for Signalling Potential Adverse Drug Reactions from Administrative Health Databases”,pg:867-876,2006.

[19] L. Geng and H.J. Hamilton, “Interestingness Measures for Data Mining: A Survey,” ACM Computing Surverys, Vol-38, No-3, Article-9, 2006.

[20] 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, 2003.

[21] Christopher C. Yang, Haodong Yang, And Ling Jiang, , “Postmarketing Drug Safety Surveillance Using Publicly Available Health-Consumer-Contributed Content in Social Media”, ACM Transactions on Management Information Systems, Vol- 5, No-1, 2014

[22] R. Agrawal and R. Sreikant, “Fast Algorithms for Mining Association Rules,” Proc. 20th Int’l Conf. Very Large Databases, 1994.

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

[24] W. Klosgen, “Explora: A Multipattern and Multistrategy Discovery Assistant,” Advances in Knowledge Discovery and Data Mining, U.M. Fayyad, et al., eds., first ed., MIT Press, pp. 249- 271, 1996.

[25] N. Lavrac, P. Flach, and B. Zupan, “Rule Evaluation Measures: A Unifying View,” Proc. Ninth Int’l Workshop Inductive Logic Programming, 1999.

[26] G.M. Weiss, “Mining with Rarity: A Unifying Framework,” ACM SIGKDD Explorations Newsletter, Vol-6, pp. 7-19, 2004.

[27] B. Liu, W. Hsu, and Y. Ma, “Mining Association Rules with Multiple Minimum Supports,” Proc. Fifth ACM SIGKDD Int’l Conf. Knowledge Discovery and Data Mining, 1999.

[28] 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.

[29] Y.-H. Hu and Y.-L. Chen, “Mining Association Rules with Multiple Minimum Supports: A New Mining Algorithm and a Support Tuning Mechanism,” Decision Support Systems, Vol- 42,pp. 1-24, 2006