www.ijecs.in
International Journal Of Engineering And Computer Science ISSN:2319-7242
Volume 3 Issue 5, May 2014, Page No. 5815-5818

Mine Adverse Drug Reaction from Patient Electronic Database using Particle Swarm Optimization

[1]M.Maheswari, [2]Ms. J.Preethi

[1]PG Scholar
Department of Computer Science and Engineering
Anna University Regional Centre, Coimbatore.
maheswari5504@gmail.com;

[2]Assistant Professor
Department of Computer Science and Engineering
Anna University Regional Centre, Coimbatore.

Abstract: *Data mining is the process of extracting or mining the knowledge from large amount of data. It is used in many real world applications. Mining the data from data set and analyzes the data for decision making process. Various data mining algorithms are applied to improve the decision making process. To discover the causal relationship from electronic patient datasets capture the causality among infrequent events. The proposed system to find the optimum membership function of a fuzzy system using particle swarm optimization. PSO has no evolution operators such as crossover and mutation. The particle swarm optimization concept consists of, at each time step, changing the velocity of each particle toward its p_{id} and p_{gd} locations. All particles are moved to the test optimum solution. These mining adverse drug reaction signals are useful to discover patient critical condition and reduce the computational complexity for drug safety professionals.*

Keywords: Adverse Drug Reaction (ADR), Association Rules, Data Mining Algorithms, Recognition Primed Decision (RPD) Model, Particle Swarm Optimization (PSO) Algorithm.

1. Introduction

The data mining step interacts with a user or a knowledge base. There are different data repositories on which mining can be performed. Frequent patterns are the patterns that occur frequently in data. Mining frequent patterns lead to the discovery of interesting associations and correlations within data. Classification is the process of finding a model that describes and distinguishes data classes or concepts. The process of finding interesting, interpreted, useful and novel data from a large set of data is known as Knowledge Discovery in Databases (KDD). A drug used is an appropriate dose may cause one or more adverse drug reactions. That drug is used by the patient at a long period of time the patient's medical condition will be complicated to increase the various health problem like death. The adverse drug reaction signals are mined from the post marketing surveillance. Because the premarketing drugs are new one that are not approved for marketing and the premarketing drugs are also having small size only approved for marketing. This mining is discovered based on drug and Symptoms and their support count.

The maximum support count pair is taken for these mining adverse drug Reaction (ADR) signal pairs.

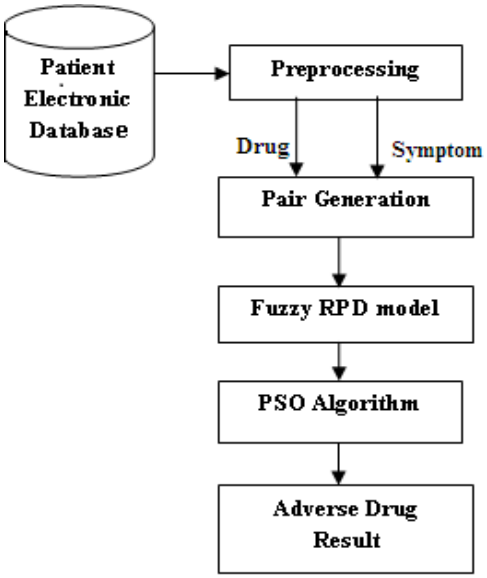


Figure 1: Mining causal association for adverse drug

Using the fuzzy recognition primed decision model is used to discover the Adverse Drug Reaction (ADR) pairs. The computational complexity is analyzed. They take maximum amount of time and the performance is less. The membership function has been tuned. All pairs are moved to the finding Adverse Drug Reactions used by the Particle Swarm Optimization (PSO). Optimized membership functions provide the better performance compare to the fuzzy model. The speed of research is very fast and the optimization of this process is improved.

To discover the causal association work summarized as follows:

- To effectively reduce the undesirable effects caused by the infrequent evens using the measure of exclusive causal leverage measure.
- The computational complexity is analyzed. Applying the Fuzzy Recognition Primed Decision (RPD) model.
- For reducing the computational complexity developed the new measure is Particle Swarm Optimization (PSO) technique.

2. Related Work

The association between two events is discovered by the area of data mining. Normally the association rules is in the form of X -> Y, where X and Y are two event sets. The interestingness measures of support and confidence are proposed for association rules [23]. Support in the association rules means X -> Y occurred at least once among the entire event sets. The association rules defined the confidence conf (X ->Y) = supp(X ->Y)/supp(X ->) in the sequence that contain X. The minsupp and minconf contains a very small set, possibly accessing false associations. The causal relationship defined as first intrinsic asymmetric causal relationship when an event X causes event Y. Second the causality is based on time dependencies like the dose has taken its caused effect is some temporal time of period. They all are discovered the infrequent patterns.

There are four different data collection and monitoring agents, each of which corresponds to a specific type of data source for Adverse Drug Reaction (ADR) detection (i.e., administration database, pharmacy database, laboratory database, and electronic patient medical records). These agents will receive data collection tasks from other agents, and then, query the corresponding databases for desired data.

ADR rank all the symptoms from the database. ADR patterns as a drug probably causing a symptom/condition can play a key role in the prevention or correction. Using the Adverse Drug Reaction (ADR) patterns, computerized systems can search health records to monitor adverse events to find patient groups at risk and to help general practitioners (GPs) ameliorate their diagnoses and prescriptions [2].
Some of the unknown Adverse Drug Reaction (ADR) in post marketing surveillance as early as possible is of great importance. The current approach to the post marketing surveillance is taken spontaneous report.

3. Existing System

3.1 Data Preprocessing

Collecting patient details from a electronic patient dataset. The two types of information need to preprocess the dataset. There are List of all drugs and List of all Symptoms and their support count. The preprocessing is avoided forming unnecessary pairs. This will reduce the computational complexity of the data mining algorithm.

3.2 Pair Generation and Evaluation

The pair will be generated from preprocessed dataset. After getting the list of drugs D, list of symptoms S all possible pairs are formed. The pairs are generated like Drugs-Symptoms pair only not interested Drug-Drug pairs and Symptoms-Symptoms pairs. Normalizing the dataset is used to reduce the computational complexity. The computational complexity is $O(m \times n)$ where m and n are the number of drugs and symptoms, respectively.

3.3 Fuzzy RPD Model

After forming the Drug-Symptom pairs <X, Y>.calculate the Degree of Causality C<X,Y> by matching the cue values defined as the experiences. The experiences are work as cues, goals, actions, expectancies. Here to categorize the degree of causality such as possible, likely, unlikely, no effect by the cue values. Actions defined what the corresponding situation is. Goal is defined what will be achieve the aim of the project. Expectancies are defined what is the next will occurred. All the cue values are extracted from the sequence by use a vector $V= (c_1,c_2,\ldots c_i,\ldots c_m)$.

4. Proposed System

4.1 Data Mining Algorithm To Mine The Causal Relationship

To mine the infrequent patterns it is used for measuring the support count from drug and symptom pairs. The support count is calculated by <d, s> pairs it is called as causal leverage value.

Drug	Support Count
d_1	σ_1
d_2	σ_2
…	…
…	…
d_m	σ_m

(a)

Symptom	Support Count
s_1	σ_1'
s_2	σ_2'
…	…
…	…
s_n	σ_n'

(b)

Figure 2: (a)Drug Hash Table(b)Symptom Hash Table

The reverse pairs <s, d> and the support count are evaluated as reverse causal leverage value. The Exclusive casual leverage value is calculated by subtracting the reverse causal leverage values from its normal casual leverage value. The minimum support count is used to further reduce the number of pairs. The maximum support count pairs are evaluating the causal relationships.

4.2 Find Adverse Drug Reaction

The causal relationships of drug and symptom pairs will be extracted from possible set of pair values. These pairs are compared to the cue values V in the Experience Knowledge Base (EKB). The similarity values are then normalized. The normalization is used to compute the causality C<X, Y> of the current patient condition. If the degree of causality C<X, Y> is greater than 0 it is added to the accumulated votes and the cases is increased by 1. Computing all the pairs of cases and it is ranked in decreasing order based on the cue values. The ranking decreasing ordered pair is according to the exclusive causal leverage values. The higher order value is exists in the causal association relationship.

4.3 Particle Swarm Optimization

PSO is a computational method that robust optimization technique is based on moving to the best solution. PSO applies the concept of social interaction to solve the problem. Each particle is treated as a point in N-dimensional space. PSO is trying to improve a candidate solution with regard to a given measure of quality. Each particle is moved to the best known local position while collision will occurred and it is moved to the search space. Each particle has the different velocity. These particle dimensions represent a fuzzy membership function parameter value. This columns represents an input and output variables such as number. Complete rule-base was considered. A rule considered complete when all possible combinations of input membership functions of all the input variables participate in fuzzy rule-base formation. The global best position marked by represents the position yielding the lowest error amongst all the P_{gd}

During the iteration every particle in the swarm is updated using the following two equations:

$$V_{id}(t+1) = W.V_{id}(t) + c_1 r_1 (p_{id} - X_{id}(t)) + c_2 r_2 \left(p_{gd} - X_{id}(t)\right) \quad (1)$$

$$X_{id}(t+1) = X_{id}(t) + V_{id} \quad (2)$$

Where,

$V_{id}(t+1)$ and V_{id} (t) are the updated and current particles velocities.

X_{id} (t + 1) and X_{id} (t)the updated and current particles positions.

C_1 and C_2 are two positive constants and $r_1$and $r_2$random numbers within the range [0,1].

Algorithm of PSO

Initialize population

Do

For k = 1 to Maximum Iteration Number K

For i = 1 to Population Size N

Update the personal best position

Update the global best position

For d = 1 to Dimension D

Update the velocity id v

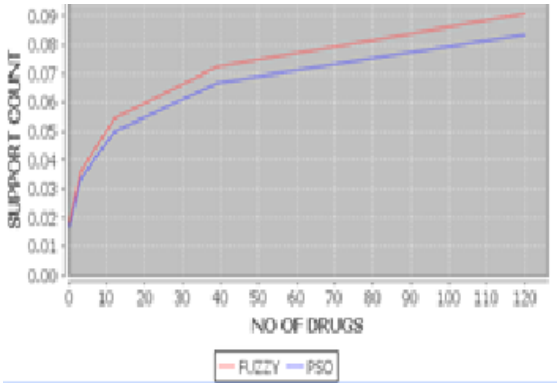
Update the position id x

Next d

Next i

Next k

Until termination criterion is met.



Each fuzzy membership functions has the parameter of mean and standard deviation. It starts with an initial set of parameters. After the parameters had been adjusted using optimization method, this parameter will be used to check the performance of the fuzzy logic. This process is repeated until the goal is achieved.

5. Developed A New Measure Foe Finding Adverse Adr Signal Pairs

To analyze the patient database if the drug-symptom pairs are already occurred or not is based on the patient id (PID). If it is already occurred means it is updated for the hash table and their support count is increased by one. Otherwise it is avoided from the forming unnecessary signal pairs.

5.1 Exclusive causal leverage measure

The degree of casualty is neared by using the exclusive causal leverage measure. The degree of causality is calculated by matching the cue values. The vector V= (c_1,c_2,…,c_i,…,c_m) to represent the cue values extracted from the sequences. Where i is the cue values, m is the total number of cues. Pair of cue values c_i and c_i' is to find the local similarity SL (c_i, c_i'). The similarity is calculated depend on the cue values.

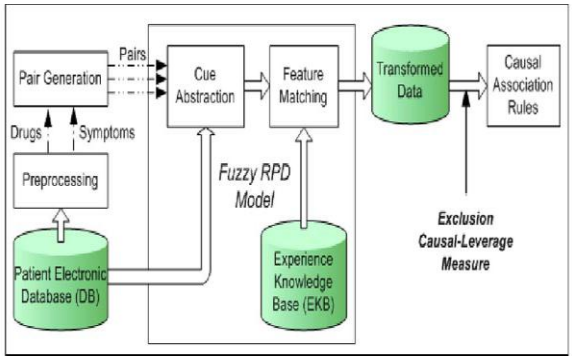


Figure 3: Mining Causal Association Rules

The global similarity is matched by the most matching the experiences with the cue values. The particular pair value from electronic patient database is abstracted from the patient cases using fuzzy rules. Also find the degree of causality at complex position of finding ADR. The temporal association is used to find ADR. Temporal association means the dose taken particular period of time taken and it causes some of the diseases for the patient.

Figure 4: Performance comparison of fuzzy and PSO

The figure 4 shows the number of drugs in x-axis and their support count in y-axis. While compare the performance between Fuzzy and PSO the performance of Fuzzy is less than the PSO as well as the accuracy, computation time, memory Fuzzy.

6. Conclusion

Mining the casual association can help us to discover the casuality of type of events and avoid its potential adverse effect. Mining these associations is very difficult when the events occur infrequently. In existing used a exclusive causal leverage measure based on Fuzzy Recognition Primed Decision model (RPD). Our new measure is particle swarm Optimization (PSO) is better than the fuzzy RPD model. The problem of Fuzzy RPD model is lacking analytical technique and selects the membership function. When occurred infrequent patterns these mining associations are very difficult. The data mining algorithm was developed and searches the potential ADR. To overcome this problem need to tuning the membership function with minimizing the error rate and maximize the performance. The experiment result showed adverse drug reaction signals rank high for all the symptoms in the patient electronic database.

References

[1] G Yanqing Ji, Hao Ying, Fellow, IEEE, John Tran, Peter Dews,Ayman Mansour, and R. Michael Massanari "A Method for Mining Infrequent Causal Associations and Its Application in Finding Adverse Drug Reaction Signal Pairs"VOL. 25, NO. 4, APRIL 2013.

[2] 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.

[3] 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.

[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] 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.

[6] Y. Ji, H. Ying, J. Yen, and R.M. Massanari, "A Fuzzy Logic-Based Computational Recognition-Primed Decision Model," InformationScience, vol. 77, pp. 4338-4353, 2007.

[7] J. Talbot and P. Waller, Stephens' Detection of New Adverse DrugReactions, fifth ed. John Wiley & Sons, 2004.

[8] L.T. Kohn, J.M. Corrigan, and M.S. Donaldson, To Error is Human:Building a Safer Health System. Nat'l Academy Press, 2000.

[9] 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.

[10] 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.