INTERNATIONAL JOURNAL OF RESEARCH IN COMPUTER APPLICATIONS AND ROBOTICS
ISSN 2320-7345

A NOVEL ICU CLINICAL DECISION SUPPORT SYSTEM USING TEMPORAL ASSOCIATION RULE MINING

R.Poornima [1], R.Karthikeyan [2]

[1] *PG Scholar, PSNA College of Engineering and Technology, Dindigul, poorni11.01@gmail.com*
[2] *Associate Professor, PSNA College of Engineering and Technology, Dindigul,karthikeyan@psnacet.edu.in*

Abstract

The modern ICU generates large volumes of complex and multimodal data. Interpreting and utilizing this information is challenging for the ICU Physician. By enhancing the ICU Clinical Decision Support System its outcomes in critical patients will be improved by providing real-time decision support, decreasing medical errors, and minimizing life-threatening events caused by delayed or uninformed medical decisions. In addition to basic patient demographics, pre-existing co morbidity data, and medication usage, it also emulate more event categories such as nurse-verified chart events, laboratory tests, and fluid balance records. For making critical decisions, these events based categories with event duration are more helpful. For handling this event duration, we are proposing the temporal association rule mining in our proposed system. From the experimentation, we can say that our proposed system has higher accuracy rate compared to the existing system. In addition to that our proposed system is good to assist ICU clinicians in making critical decisions.

Keywords: Intensive Care Units(ICUs),Association Rule Mining (ARM)

1. Introduction

According to the Society of Critical Care Medicine (SCCM), there are approximately five million patients admitted annually to intensive care units (ICUs) in the United States. Average mortality rates ranging from 10% to 29% [1], which are the highest rates of all the units in a hospital. Compared to clinical settings, the ICU has some of the highest rates of medical errors [2, 3]. With the extensive hemodynamic monitoring and use of multiple measurement technologies, the modern ICU generates large volumes of complex and multimodal data. Interpreting and utilizing this information is challenging for the ICU physician.

We have designed and developed an ICU clinical decision support system (CDSS) to improve outcomes in critically ill patients by providing real-time decision support, decreasing medical errors, and minimizing life-threatening events caused by delayed or uninformed medical decisions. CDSSs are computer-aided ``active knowledge systems which use two or more items of patient data to generate case-specific advice'' [4] and it can improve a physician's decision making performance for providing an evidence strongly [4] . For optimal medical decision making, the CDSS needs to be data-driven, rapid, and informed.

Evidence-based medicine is the ``conscientious, explicit and judicious use of current best evidence in making decisions about the care of individual patients'' [5]. A CDSS is evidence-based if its knowledge base is derived from, and continually reflects, the most up-to-date evidence from the practice-based sources [6]. The IF-THEN rule is a generic form of evidence in an evidence-based CDSS. The rule implies that IF an antecedent (i.e., a set of conditions) presents, THEN an outcome is an action should be taken. Evidence-based CDSSs have provided important risk assessment scores for clinicians, such as prediction of ICU survivability (e.g., APACHE II [7]), length of stay[8] , organ failure (e.g., SOFA[9]), neurologic prognosis (e.g., Glasgow Coma Score[10]), and outcomes after acute coronary syndrome (e.g., GRACE ACS model[11]). However, none of the above applications are true CDSSs because they are not interactive nor flexible, which are two key features of true decision support systems[12].These computer-based ICU assistance systems claim to be CDSSs because they provide the ``statistics'' of evidence. However, it is difficult for clinicians to make a correct decision by recalling all corresponding knowledge based on these statistics in a timely fashion. They have to search their archives, find the relevant evidence (assuming it is up-to-date). The luxury of time is rare, therefore this decision support process is not feasible in the critical care setting. Thus, a reliable CDSS in an ICU should provide not only statistically significant knowledge, but also an interactive user interface that enables clinicians to search for evidence effectively and in real-time.

Besides being interactive, an evidence-based ICU CDSS need to be flexible. In typical ICU systems that claim to be CDSSs, researchers define expected IF-THEN rules given certain clinical problems, validate the rules, and form new evidence if the rule is statistically significant. The clinical conditions are matched, and then the clinicians can refer to the evidence. For example, a clinician can select an appropriate antimicrobial drug for a septic patient when the pathogenic organism has specific hemodynamic and biochemical markers. However, some patients lack clear-cut evidence for the presence of an infection and/or the type of infecting organism, which makes the decision to treat with an antimicrobial drug experience-based instead of evidence based.

The clinician still needs to make the same decision about antimicrobial drug prescription with incomplete information, and then passively assess the diagnosis. Such a process introduces human bias that deviates from the original design of the CDSS. Clinicians face this challenge on a daily basis for every patient in the ICU due to heterogeneous conditions. Therefore, a flexible ICU CDSS is needed to allow clinicians to customize conditions to better describe a patient's immediate status (i.e., personalized), instead of referring to fixed evidence formed from different clinical situations.

A powerful CDSS relies on a sufficient and representative database of patient ICU stays. Although the bedside monitor can generate large amounts of data from each patient as compared to other care settings [13], the number of unique patients in a standard ICU CDSS database is typically small. Thus, the accuracy of new decision support evidence is limited by the diversity of phenotypes contained in a small number of patients. Even though some studies contain large numbers of patients, the evidence is usually mined from the entire cohort without clinical categorization. Ideally, before applying data modelling analysis, a CDSS should first extract a cohort of patients who have similar medical histories and situations, and reveal the sample size as a reference when delivering new evidence back to clinicians. Clinicians can judge when the evidence is mined from a sufficient and representative dataset.

To our knowledge, to date, no true CDSSs have been developed for the ICU. To address the aforementioned challenges, we designed and developed an ICU CDSS called icuARM. To improve the performance of the system, we proposed the mining approach called temporal association rule mining. The report of ICU Clinical Decision Support system as the following structure. First, provide the data source. Then, we describe the detailed association rule mining approach. Next, we generate the temporal association rule mining adds time constraints on association rule.

2. Background and related work

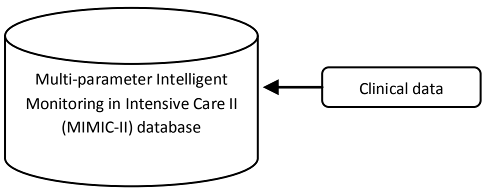
Data Source is selected from MIMIC II database, described as SPECT heart data. Association rule mining raised by Rakesh Agrawal is an important research problem in data mining field. Association rule mining aims at detecting the relationship of tuples in transactional database and serving decision making (Rakesh Agrawal et al, 1993). Let's data itemset X = {X1, X2,..., Xm}, Y = {Y1, Y2, ..., Yn}, association rule can be

represented as the form of "X→Y". Association rule X→Y expresses "tuples satisfying conditions in X also satisfy conditions in Y.
If rule X→Y is true, then it has the support and confidence values. That is, association rule mining is to find the rule sets satisfying minimum support threshold and minimum confidence threshold in the dataset or database.
Temporal association rule from "Temporal Association Rule on T-Apriori Algorithm and its typical application". Comparing with association rule mining, temporal association rule adds time constraint (it can be time point or time range) on association rule. A transaction with time information can be described as: {TID, I1, I2 …In, Ts, Te}. TID is the ID for each transaction; n-itemsets means there are n items in the itemset; Ts and Te represent the start and the end of valid time respectively.Valid time means the event occurring time, while transaction time the database time. Ts may equal Te, such as sale records in the supermarket (the transaction occurs at one moment). We can give the definition of temporal association rule as "Let min_s and min_c represent minimum support threshold and minimum confidence threshold respectively, if and only if during [ts, te], support ≥ min_s, confidence ≥ min_c, rule X→Y is a temporal association rule, which could be described as X→Y (support, confidence, [ts, te]). Itemset means the collection of items. If there were k items, we call it k-itemsets. The itemsets that satisfies min_s is called frequent itemset"
Introduces the problem of mining that is, a large collection of basket data type transactions for association rules between sets of items with some minimum specified confidence, and presents an efficient algorithm for this purpose. Another as pruning functions can prune out itemsets as soon as they are generated. To select the optimal threshold in a structured fashion used in clinical practice. Here used this method to dichotomize the outcome length of stay at the Intensive Care Unit (ICU LOS). To select the threshold value for dichotomization of survival outcomes, such as ICU LOS, in a structured fashion [8].

3. Methods and Procedures

3.1 Data source - the MIMIC-II data:

The data in the icuARM is imported from the Multi-parameter Intelligent Monitoring in Intensive Care II (MIMIC-II) database. MIMIC-II is a publicly accessible ICU data repository containing records of over 40,000 ICU stays. The data in MIMIC-II can be categorized into two major categories: clinical data and physiological data. The data mining process only includes clinical data the clinical data is collected from MIMIC-II's ICU information systems and hospital electronic health record systems. In our proposed system, we generate temporal data mining.
The imported ICU data set consists of a sample of 200 subjects .We further divide the MIMIC-II clinical data into two groups of categories: basic and event-based. The basic categories include data that remain unchanged during one ICU stay (e.g., patient demographics and pre-existing comorbidities). The event-based categories contain data collected at multiple time points within an ICU stay, including laboratory tests (e.g., blood chemistries, complete blood counts), medication events (e.g., insulin, heparin), fluid balance (e.g., urine output), and nurse-verified chart measurements (e.g., blood pressure, heart rate). During an ICU stay if an event occurs in a sequential order for a particular patient, then the values in event-based categories as either 0 or 1 that can be stored in database. Durations of chart measurement, medication, and fluid balance events are also imported.



3.2 Frequent itemset generation:

To generate a frequent itemset, in order to discover frequent and confident association rules, the mining process requires users to specify a minimum values as thresholds to drop infrequent rule, which is minimum support ($Supp_{min}$). By using minimum support ($Supp_{min}$) to prune out the infrequent items. Rules are considered to be frequent if their supports are at least $Supp_{min}$. The goal of ARM is to find all frequent rules based on these two users specified values.

There are two main steps in revealing association rules. The first step is to find all frequent itemsets that have supports above $Supp_{min}$. The second step is to use the frequent itemsets to generate confident rules with confidences above the $Conf_{min}$. The Apriori algorithm is the most popular in ARM research.

The following algorithm is as follows:

Algorithm: F = Apriori(T,I,$Supp_{min}$)
Input: T(transactions), I(1-itemsets), $Supp_{min}$
Output: F (Frequent Itemsets)
F_1= {f|I,f.support>=$Supp_{min}$};
for(k=2;F_{k-1} NOT EQUAL to ϕ; k++) do
    C_k = GenCandidate(F_{k-1});
    for each transaction t ϵ T do
        for each candidate c ϵ C_k do
    if c is contained in t then
    c.count++;
    end
end
F_k = { c ϵ C_k | c.support >= $Supp_{min}$}
end
Return F= U_k F_k;

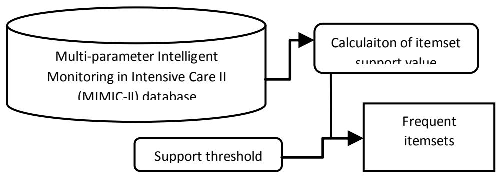
The Apriori algorithm utilizes an iterative process to generate frequent itemsets. Let I = {I1, I2, . . . , IN } consist of N possible items in the database. In the first iteration, the algorithm starts by counting the occurrence of 1-itemset candidates that contain only one item. 1-itemset candidates that have supports lower than $Supp_{min}$ are pruned out and the remaining ones are called frequent 1-itemsets. In the following iterations (i.e., k >1), the candidate k-itemsets are first generated by joining the frequent (k-1)-itemsets. Then frequent k-itemsets are generated by pruning out candidate k- itemsets that have supports lower than $Supp_{min}$. The iteration continues until no more candidates or frequent itemsets can be found.

The GenCandidate in the Apriori algorithm is the candidate itemset generation algorithm that is given as follows:

Algorithm: C_k = GenCandidate (F_{k-1})

Input: F_{k-1} (Frequent k-1 itemsets)
Output: C_k(Candidate k itemsets)
$C_k = \phi$;
forall $f_m, f_n \in F_{k-1}$
    where $f_m = \{i_1, \ldots, i_{k-2}, i_{k-1}\}$
    and $f_n = \{i_1, \ldots, i_{k-2}, i'_{k-1}\}$
    and i_{k-1} NOT EQUAL i'_{k-1} do
C = $\{i_1, \ldots, i_{k-2}, i'_{k-1}\}$;
$C_k = C_k$ U {c};
foreach (k-1)-subset s of c do
    if(s NOT $\in F_{k-1}$) then
        delete c from C_k;
  end
end
Return C_k;

The generation of frequent itemset represented as Ck is a superset of Lk using minimum support count.

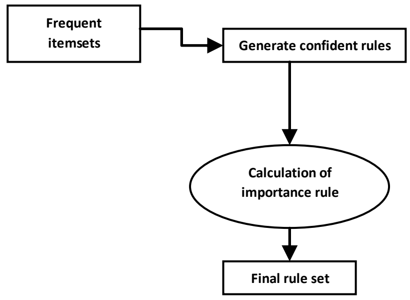


3.3 Building Association rules:

After generating all frequent itemsets via the Apriori algorithm, the second subproblem is to generate confident rules that satisfy $Conf_{min}$. For each frequent itemset f, consider all non-empty subsets of f . For each subset a, the process forms new rule (f - a) if its confidence is above $Conf_{min}$. We can call a and (f - a) the antecedent and consequent, which are the X and Y , respectively, of an association rule.If a rule $X \Rightarrow Y$ has a high level of association according to its confidence. However, a high confidence of rule $X \Rightarrow Y$ still cannot guarantee a low confidence of its counter case. That is, during clinical decision making, we want to consider when the rule $X \Rightarrow Y$ yields a higher confidence than its counter case. This means that Y is likely to occur only when X occurs, and when X does not occur, Y has a low chance of occurrence. We use the following equation to evaluate the importance of a rule X=>Y:

Impo(X $\Rightarrow$ Y) = Conf(X=>Y) / Conf(¬X=>Y)

The importance metric ranges from 0 to1. A rule of importance less than 1 means that the antecedent predicts the consequent worse than the counter case of the antecedent. This type of rule should be ignored. Thus the rules are expected to have an importance >1. In this, the qualitatively chosen threshold of importance is selected to be >1.By setting the threshold value, to display the higher confidence rules. Rules are prune out when it does not satisfy the threshold limit, then remaining confident rules are displayed.



3.4 Generation of Frequent Itemsets Using T-Apriori algorithm:

In T-Apriori algorithm, the process of frequent itemsets generation is similar to Apriori algorithm, but we need special treatment of time information. Here T is temporal database, Lk is the frequent itemset.

Input: T, min_s
Output: Results = U_{LK} k

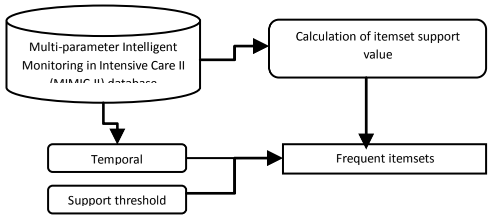
Algorithm Process:
For all RecordSets do
    Subtract (RecordSets)
end
ItemSets=TRecordSets without TRecordSets.time
C_1 ={Candidate 1-ItemSets},L_1 = { c ϵ C_1 | c.count>=min_s_1 }
for(k=2;L_{k-1} Not Equal to ϕ,k++) do begin
    C_1 = Apriori_Gen (L_{k-1})
    for all transaction t ϵ ItemSets do begin
        C_1 = subset (C_1,t)
        for all Candidate c ϵ C_1 do
        c.count++
    end
    L_k={c ϵ C_k|c.count>=min_s}
    end
Results=U_kL_k

In this algorithm, represent the support count of frequent itemset and c. Apriori_Gen function generates frequent itemset:Ck. The input parameter of Apriori_Gen is Lk–1, (k-1)-itemset and the output is Ck. The generation of Ck can be divided into two steps: Join and Prune; after the generation of Ck, we need to scan the database and calculate the support of each subset of Ck. subset performs this function.

Subtract function: The input parameters are RecordSets and time threshold, it returns TRecordsets satisfying time threshold.

Input: RecordSets, [min_ts, min_te,min_t]
Output: TRecordSets
Algorithm Process:
    for all RecordSets do
        if RecordSets.time ϵ [min_t_s , min_t_e]
        or RecordSets.time ϵ [min_t] then
    TRecordSets=RecordSets
    End

Vol.2 Issue.11, Pg.: 99-108
November 2014



Comparing with the normal association rule mining, temporal association rule adds time constraint (it can be time point or time range) on association rule. A transaction with time information can be described as: {TID, I1, I2 …In, Ts, Te}. TID is the ID for each transaction; n-itemsets means there are n items in the itemset; Ts and Te represent the start and the end of valid time respectively. Valid time means the event occurring time, while transaction time the database time. Definition of strong association rule "association rule strictly satisfies minimum support threshold and minimum confidence threshold".

3.5 Building Temporal Association Rule:

After the T-Apriori algorithm, we will focus on the algorithm for temporal association rule mining, Apriori algorithm is a basic one for getting frequent itemsets as well as an influential one in association rule mining. Through generating frequent itemsets with time information like temporal constraint, cycle and trend characteristics, we propose a temporal association rule mining algorithm: T-Apriori as to temporal dataset or database.

According to Definition of strong association rule, temporal association rule can be described as "X→Y(support, confidence, [ts,te])". T-Apriori algorithm refers time as a constraint. First of all, we need analyze the temporal database with respect to time threshold. Time threshold is the time point or time range. Time range can be expressed as [min_ts, min_te], while time point [min_t]. Then we should delete the time information in the temporal dataset or database in order to decrease computational complexity and apply the T-Apriori algorithm to generate frequent itemsets and corresponding temporal association rules.

Input: $\bigcup_{k} L_K$,, min_c

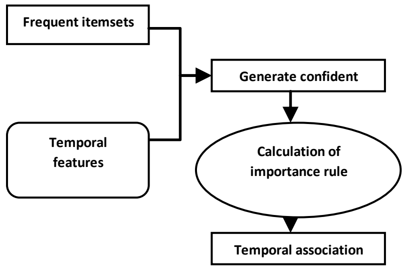
Output: rules like x→y(support, confidence, [ts, te]) or x→y(support, confidence, min_t)
Algorithm Process:

```
for all Lk,k>=2 do begin
        H1={consequents of rules derived from Lk with one item in the conquent};
        Call op-genrules {lk,H1};
        Call Add(Lk)

end
```

Add: The input parameter is frequent itemset Lk (without time information), it needs scan the TrecordSets and returns the temporal association rule.

```
for all TRecordSets(i), i>=0 do
        for all L'k (k), k>=2 do
                if L' k ϵTRecordSets(i) then
                        TRecordSets (i).time INTERSECTION of (Lk-hm+1) → hm+1
                        with confidence = conf and support=support(Lk)
        end
end
```



4. SYSTEM IMPLEMENTATION

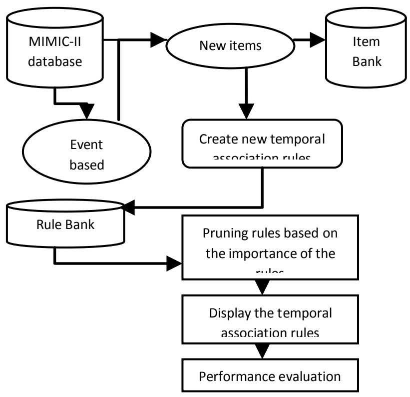
4.1 System Architecture:

In our proposed system, we will emulate more categories such as nurse-verified chart events, laboratory tests, and fluid balance records etc. to better assist ICU clinicians in making critical decisions. In order to improve the performance of the system, we are using these event based categories from the MIMIC-II data. In deeply, Nurse-verified chart events are such as heart rate, heart rhythm, blood pressure, non-invasive blood pressure, central venous pressure, respiratory rate, etc. fluid balance are urine out, free water bolus, gastric, nasogastric, etc. also laboratory tests includes the blood chemistries, complete blood counts, pH of blood, haemoglobin in blood, etc. in our proposed system, the efficiency of the system is improved.

Comparing with association rule mining, temporal association rule adds time constraint (it can be time point or time range) on association rule. A transaction with time information can be described as: {TID, I1, I2 …In, Ts, Te}. TID is the ID for each transaction; n-itemsets means there are n items in the itemset; Ts and Te represent the start and the end of valid time respectively (or the start and the end of transaction). Valid time means the event occurring time, while transaction time the database time. Ts may equal Te, such as sale records in the supermarket (the transaction occurs at one moment).

Definition of Strong Association rule mining: Let min_s and min_c represent minimum support threshold and minimum confidence threshold respectively, which could be described as X→Y (support, confidence, [ts, te]).

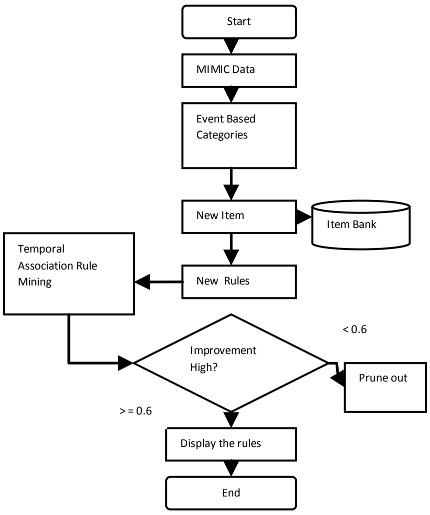
Next we will focus on the algorithm for temporal association rule mining, Apriori algorithm is a basic one for getting frequent itemsets as well as an influential one in association rule mining. Through generating frequent itemsets with time information like temporal constraint, we propose a temporal association rule mining algorithm: T-Apriori as to temporal dataset or database.



T-Apriori algorithm refers time as a constraint. First, we need analyze the temporal database with respect to time threshold. Time threshold is the time point or time range. Time ranges from min_ts to min_te, while time point [min_t], In order to decrease computational complexity and apply the T-Apriori algorithm to generate frequent itemsets and corresponding temporal association rules.

4.2 System Flow Diagram :

System flow Diagram represents the flow of events. If an event occurs that can be stored in Item Bank and generate the rule called "Temporal Association Rule".
For higher confidence we have to determine the importance rule for the threshold value set as 60%.
That will helpful to display the rules above or equal to 60%, others are prune out.



5. Conclusion and Future work

In order to improve the performance of the system, we are proposing the enhanced Clinical Decision Support system. In our proposed system, we will emulate more event categories such as nurse-verified chart events, laboratory tests, and fluid balance records etc. to better assist ICU clinicians in making critical decisions. We are using these event based categories from the MIMIC-II data. Generally, event based categories are also includes the event duration. Since, for making critical decisions these events based categories with event duration are more helpful. For handling this event duration, we are proposing the temporal association rule mining in our proposed system. In our proposed system, we have introduced time in the problem of association rules discovery, given place to what we call Temporal Association Rules. To generate association rules without time, because it adds time information on frequent itemsets. So the association rules are temporal ones. Each item and rule has now an associated lifespan, which comes from the explicitly defined time in database transactions. A solution is that the user may say which dates are old enough, so the rules with lifespan previous to those dates would be considered obsolete and not presented to the user. Furthermore, if the algorithm used to generate the frequent itemsets finds old items or itemsets, it may eliminate them directly, which it would be an additional pruning.
Our further plans to improve this CDSS in three directions. First, we aim to include continuous physiological data with corresponding time stamps from the MIMIC-II database to perform temporal association rule mining. Second, the current mining process with the Apriori algorithm requires clinicians to manually specify variables of interest and cut-points (for numerical variables) in the items of antecedents and consequents. We will develop automatic feature selection, such as the supervised mRMR method or the unsupervised MCFS method, and discretization for more objective item construction.

Table 1 Comparison of Algorithms

Algorithm	Constraint Dependence	CDSS Performances
Association Rule Mining	Present	Low
Temporal Association Rule Mining	Present and future	Medium
Supervised mRMR method	Present ,future and predicted values	High

The following Table 1 shows the advantage of Bayesian estimator over the existing estimators. The present estimator i.e., Bayesian estimator provides better results by combining advantages of two other estimators. Bayesian estimator combines both advantages of ML and MaP also includes predicted values. These predicted values compares with previous results for better results. In future, the implementation of the project can done using GPU in order to achieve real time performance tracking of many subjects. The speedup is achieved using benchmarks.

REFERENCES

[1] Kurgan, L.A., Cios, K.J., Tadeusiewicz, R., Ogiela, M. & Goodenday, L.S."Knowledge Discovery Approach to Automated Cardiac SPECT Diagnosis"Artificial Intelligence in Medicine, vol. 23:2, pp 149-169, Oct 2001.

[2] ZHAI Lianga, TANG Xinmingb, LI Lina,JIANG Wenlianga,"TEMPORAL ASSOCIATION RULE MINING BASED ON T-APRIORI ALGORITHM ".

[3] D. J. Cullen, B. J. Sweitzer, D. W. Bates, E. Burdick, A. Edmondson, and L. L. Leape, ``Preventable adverse drug events in hospitalized patients: A comparative study of intensive care and general care units," Critical Care Med., vol. 25, no. 8, pp. 1289_1297, 1997.

[4] L. B. Andrews, C. Stocking, T. Krizek, L. Gottlieb, C. Krizek, T. Vargish, et al., ``An alternative strategy for studying adverse events in medical care," Lancet, vol. 349, pp. 309_313, Feb. 1997.

[5] J.Wyatt and D. Spiegelhalter, ``Field trials of medical decision-aids: Potential problems and solutions," in Proc. Annu. Symp. Comput. Appl. Med. Care, 1991, pp. 3_7.

[6] I. Sim, P. Gorman, R. A. Greenes, R. B. Haynes, B. Kaplan, H. Lehmann, et al., ``Clinical decision support systems for the practice of evidence-based medicine," J. Amer. Med. Inf. Assoc., vol. 8, no. 6, pp. 527_534, 2001.

[7] R. Agrawal, T. Imielinski, and A. Swami, ``Mining association rules between sets of items in large databases," in Proc. ACM SIGMOD Rec., 1993, pp. 207_216.

A Brief Author Biography

***R.Poornima* –** PG Scholar at PSNA College of Engineering and Technology, Dindigul, India. Research interest is Data Mining.

***R.Karthikeyan* –** Associate Professor, PSNA College of Engineering and Technology, Dindigul, India. Research interest is Data Mining.