ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 4, April 2015

Clinic + - A Clinical Decision Support System Using Association Rule Mining

Sangeetha Santhosh, Mercelin Francis

M.Tech Student, Dept. of CSE., Marian Engineering College, Kerala University, Trivandrum, Kerala, India

Asst. Professor, Dept. of CSE., Marian Engineering College, Kerala University, Trivandrum, Kerala, India

ABSTRACT: A Clinical Decision Support System(CDSS) assist practitioners and healthcare providers in decision making through timely access to electronically stored medical knowledge. A CDSS interacts with practitioners and electronic medical record systems to receive the patient data as input and provides reminders, alerts, or recommendations for patient diagnosis, treatment and long-term care planning. A review on this clinical decision support system , have observed an ICU clinical decision support system based on associate rule mining (ARM), and a publicly available research database MIMIC-II (Multi-parameter Intelligent Monitoring in Intensive Care II) .System is focused on developing clinical decision support system considering both clinical and physiological data using association rule data mining algorithm and clustering algorithm.

KEYWORDS: Clinical Decision Support System, Association Rule Mining, Apriori Algorithm, Clustering Algorithm, Electronic health Record.

I. INTRODUCTION

Health information technology, especially EHR, can allow clinicians to have a real-time access to complete patient data, the involved systems help them to make accurate clinical decisions. To make an accurate decision, clinicians must consider EHRs, clinical databases and clinical knowledge bases [1]. EHRs are distributed and are very large; they contain the complete history of patient health data. Clinicians must collect, compare and analyze data from these sources and make a timed decision. This information overflow could cause physicians to disregard vital information (EHR data, hidden knowledge in clinical databases and EHRs, and knowledge bases) and could make it take a long time to make a correct decision. Therefore, they need an automated system that helps in collecting, calculating and analyzing all of the available data and helps in making decisions. This type of system is called a CDSS.

CDSSs are interactive computer programs that are designed to assist physicians and other health professionals. They help in drug prescriptions, diagnosis and disease management, to improve services and reduce costs, risks and errors. The CDSS can check for patient drug allergies, compare drug and laboratory values, evaluate the potential for drug–drug interactions, suggest drug alternatives, block duplicate orders, suggest drug doses, and frequencies and provide recommendations. In addition, a CDSS can provide clinical knowledge and best practice standards and guidelines for non-expert physicians. A CDSS provides recommendations that are based on the available patient-specific data (EHR) and medical facts (knowledge bases). The EHR is continuously updated; thus, the knowledge bases must be continuously updated by discovered knowledge from domain experts and discovered knowledge from EHRs and clinical databases. Many applications in biomedical informatics require clinical knowledge bases, relating signs and symptoms to diseases screening tests and indications or diseases and medications. These knowledge bases are often developed and maintained by experts, at significant cost. However, automated methods for developing such knowledge bases hold promise. There are a set of data mining techniques which can be used to automatically infer relationships between medications, laboratory results and problems, and validate a knowledge base.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 4, April 2015

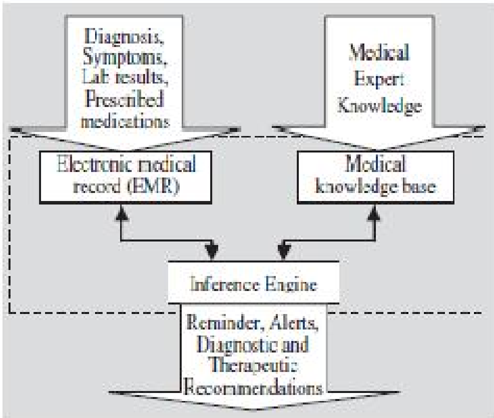

Fig: Model of a Decision Support System

II. RELATED WORK

In [1] authors developed a real-time clinical decision support system icuARM to assist clinicians in generating quantitative and real-time decision support rules for the ICU based on a large ICU patient database MIMICII. Paper adopted the' support' and the 'confidence' metrics suitable for ICU clinical application from conventional association rule mining.

Paper [2] provides technical architecture takes advantage of Electronic Health Record (EHR), data mining techniques, clinical databases, domain expert knowledge bases, available technologies and standards to provide decision-making support for healthcare professionals. The architecture will work extremely well in distributed EHR environments in which each hospital has its own local EHR, interoperability and scalability objectives of an EHR. The system will also have a set of distributed knowledge bases. Each knowledge base will be specialized in a specific domain (i.e., heart disease), and the model achieves cooperation, integration and interoperability between these knowledge bases. Model will solve the problem of having multiple healthcare providers; each provider has its local and large EHRs, knowledge base, and clinical databases.

[8][9]Within Temporal Data Mining, research focuses on the analysis of time series, collected measuring clinical or biological variables at different points in time. One of the most attractive applications of TDM concerns the extraction of temporal rules from data[10]. Unlike association rules, temporal rules are characterized by the fact that the consequent is related to the antecedent of the rule by some kind of temporal relationship; moreover, a temporal rule typically suggests a cause-effect association between the antecedent and the consequent of the rule itself. Possible associations between those patterns and the non-adherence events are mined with an APRIORI-like procedure. Works are mainly dealing with providing decision support, using physiological data and efficient grouping of multiple data is lacking.

III. PRINCIPLE OF ASSOCIATION RULE MINING

A. *Data Source - MIMIC-II database*:

The data is imported from the Multi-parameter Intelligent Monitoring in Intensive Care II (MIMIC-II) database [2]. MIMIC-II is a publicly accessible ICU data repository containing records of over 40,000 ICU stays in which 32,000 are adult (>15 yrs) records and 8,000 are neonatal (<2 yrs) records [2]. The data in MIMIC-II can be categorized into two major categories: clinical data and physiological data. The clinical data is collected from MIMICII's ICU information systems and hospital electronic health record systems. The high-resolution physiological data consists of time series waveforms and time series measurements from bedside monitors. MIMIC-II data were also identified in order to remove protected health information. Any interested researcher can gain

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 4, April 2015

access to MIMIC-II free of charge after signing a data use agreement and completing human subjects training. MIMIC-II can support a wide variety of research studies, ranging from the development of clinical decision support algorithms to retrospective clinical studies.

B. *Association Rule Mining*:

After importing data from MIMIC-II, a sophisticated mining process is required to unearth meaningful associations from such a voluminous dataset. Association rule mining (ARM) is a method to reveal meaningful relations between variables in databases.[3] ARM has been widely adopted in applications such as heart disease prediction, healthcare auditing, and neurological diagnosis.

Rules in ARM are in the form of $X \Rightarrow Y$, which means that X implies Y , where X and Y are called antecedent and consequent, respectively. In its original marketing analysis context, the rule $X \Rightarrow Y$ carries the meaning that if a customer buys items in X, he/she is also likely to buy items in Y. In MIMIC-II, one patient may be associated with one or more hospital stays; and one hospital stay may be associated with one or more ICU stays. Therefore, the most basic data piece in our mining process is the ICU stay. Therefore, in ICU data mining process, a rule $X \Rightarrow Y$ implies that if X occurs in one ICU stay, Y is also likely to occur during the stay. The antecedent X and consequent Y are itemsets that consist of one or more item(s). An item is composed of a variable with a corresponding value or a range of values. An item can be numerical or categorical depending on the data type of the variable. Two important metrics-*support* and *confidence* quantify the frequency and level of association of a rule.

In order to discover frequent and confident association rules, the mining process requires users to specify two minimum values as thresholds to drop infrequent and unconfident rules, which are minimum support (*Suppmin*) and minimum confidence (*Confmin*). Rules are considered to be frequent if their supports are at least *Suppmin* and confident if their confidences are at least *Confmin*. The goal of ARM is to find all frequent and confident rules based on these two userspecfied values. There are two main steps in revealing association rules. The first step is to find all frequent itemsets that have supports above *Suppmin*. The second step is to use the frequent itemsets to generate confident rules with confidences above the *Confmin*. Since the first algorithm was introduced in the original report of ARM new algorithms have been proposed to improve the efficiency of the generation of frequent itemsets. Among these algorithms, the **Apriori** algorithm is the most popular in ARM research.

C. *Apriori Algorithm*:

The Apriori algorithm utilizes an iterative process to generate frequent itemsets. Let $I = \{I1, I2, \ldots, IN\}$ 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 *Suppmin* are pruned out and he 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 *Suppmin*. The iteration continues until no more candidates or frequent itemsets can be found.

i. F= Apriori(T,I,Suppmin)

```
// Input: T(Transactions),I(1-itemset),Suppmin
//Output: F(Frequent Itemsets)
F1 = {f|f ϵ I, support ≥ Suppmin};
For(k=2;Fk-1 ≠ Ø; k++) do
     Ck= GenCandiadate(Fk-1);
     for each ttransaction t C T do
       for each candidate c ϵ Ck do
          if  c is contained in t then
          c.count++;
         endif
       endfor
       Fk= {c ϵ Ck | c.support  ≥ Suppmin}
      endfor
return F = Uk Fk;
```

ii. Ck = GenCandidate(Fk-1)

```
// Input: Fk-1(Frequent k-1 itemsets)
//Output: Ck(Candidate k itemsets)
Ck = Ø;
forall fm,fn ϵ Fk-1
   Where fm = {i1, ......., ik-2,ik-1}
   and fn = {i1,......, ik-2, ik-1}
   and ik-1 ≠ i'k-1 do
   c = {i1,......, ik-1, ik-2}
   Ck = Ck U {c};
   foreach (k-1) – subsset s of c do
      if ( s !ϵ Fk-1) then
         delete c from Ck;
      endif
   endfor
return Ck;
```

IV. PROPOSED WORK

Develop a clinical decision support system to incorporate both clinical data and physiological data in data collection to improve efficiency of decision making and compare results with existing system. Relationship between dataset is calculated using **association rule mining algorithm**. Each dataset need to be grouped so that they can be efficiently mined and can also adopt only required information from each group. For this purpose, **clustering algorithm** is also incorporated which will divide similar dataset into clusters. The problem of **cold start problem** is avoided by creating new cluster for each new item.

V. SIMULATION RESULTS

FIG 1: ENTERING TEST RESULTS

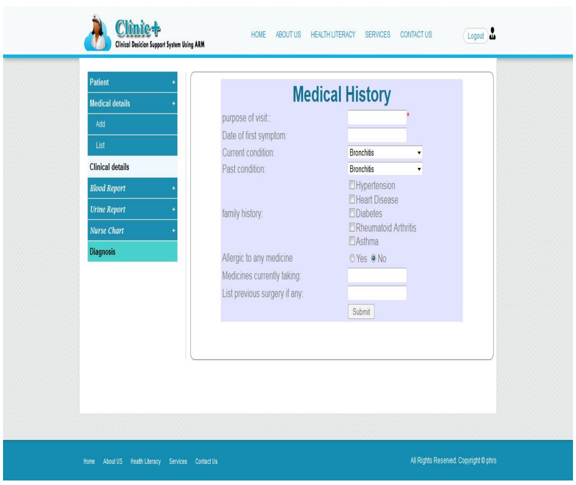

FIG 2: ENTERING MEDICAL HISTORY

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 4, April 2015

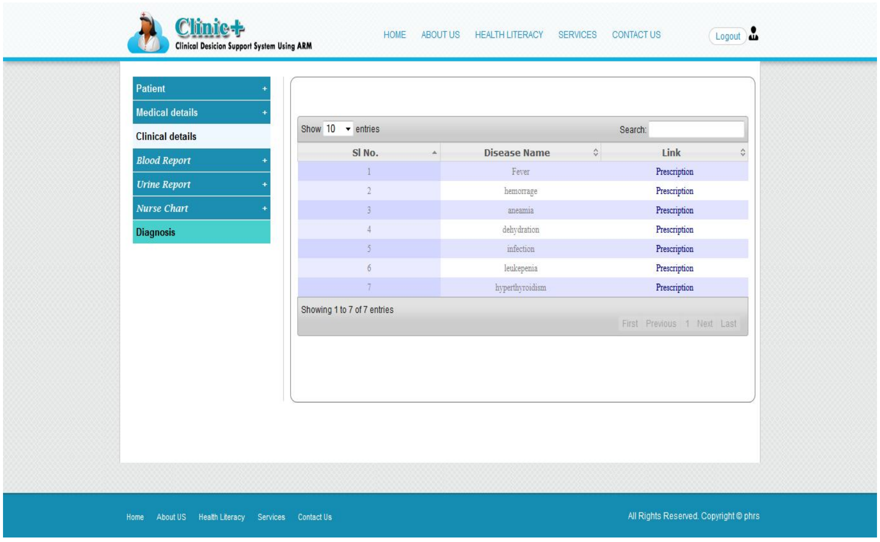

Fig 3: Listing diagnosis

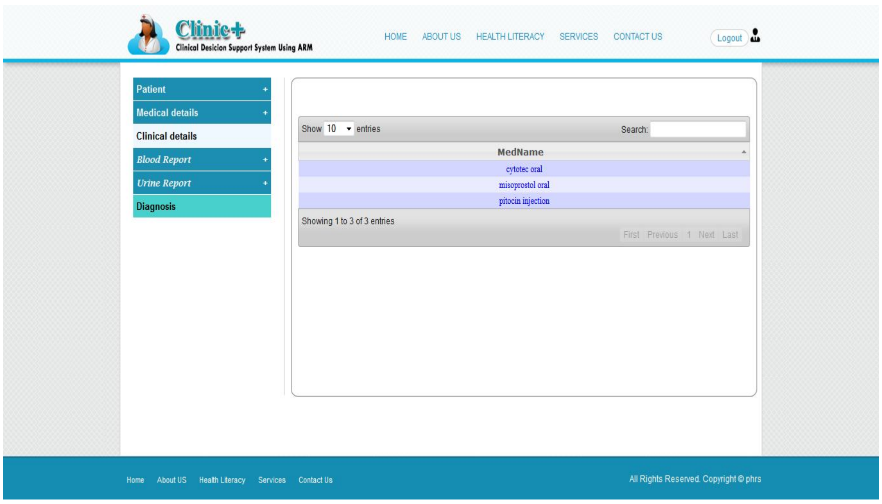

Fig 4: listing medication

System provide user friendly interface for decision support. Fig [1] is the interface for adding test results like blood test report, urine test report. Patient personal details will also be entered by staff members on their arrival. Fig [2] is the interface given for entering patient's previous medical history. After collecting all relevant information from patients like personal details, previous medical history, current patient condition and test results, appropriate suggestions of possible diagnosis is formulated by comparing patient medical history and similar cases. This process of comparing patient condition is done using efficient data mining and clustering of data. For each diagnosis suggestion there is a link which lists medication suggestions after checking previous similar cases and avoiding medications which are allergic to patient. Fig [3] is the page that displays list of diagnosis suggestions and fig [4] list of medication.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 4, April 2015

VI. CONCLUSION

Evidence-based real-time decision-making for critically ill patients in the ICU has become more challenging because the volume and complexity of the data have been increasing over the years. Thus, to assist clinicians in making optimal decisions, there is a critical need to apply modern information technology and advanced data analytics to extract information from heterogeneous clinical data. A clinical decision supporting system is designed to incorporate medical data and accessed by administrator, hospital staff and doctors. Employee management is done by admin and adding test results is done by staff. A web page will display patient test results and corresponding list of possible diseases and medications using data mining approach. Design details and system screenshots are given in this report.

REFERENCES

1. Chih-wen cheng1 (Member, IEEE), Nikhil chanani, Janani Venugopalan (Member, IEEE), Kevin Maher, and May Dongmei Wang (Senior Member, IEEE), ″icuARM _ An ICU Clinical Decision Support System Using Association Rule Mining″, IEEE journal of Translational Engineering in Health and Medicine, Volume 1, 2013
2. 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.
3. Rakesh Agrawal, Tomasz Imielinski, Arun Swami, ″ Mining Association Rules between Sets of Items in Large Databases″, IBM Almaden Research Center, Proceedings of the 1993 ACM SIGMOD Conference Washington DC, USA, May 2011
4. J. Ramon, D. Fierens, F. Güiza, G. Meyfroidt, H. Blockeel, M. Bruynooghe., ``Mining data from intensive care patients," *Adv. Eng. Informat.*, vol. 21, no. 3, pp. 243_256, 2007.
5. Shaker H, El-Sappagh , Samir El-Masri , ″A distributed clinical decision support system architecture″, Journal of King Saud University Computer and Information Sciences (2014) 26, 69–78
6. Adam Wright , Elizabeth S. Chen , Francine L. Maloney e, ″An automated technique for identifying associations between medications, laboratory results and problems″, Journal of Biomedical Informatics 43 (2010) 891–901
7. M. Saeed, M. Villarroel, A. T. Reisner, G. Clifford, L.-W. Lehman, G. Moody, *et al.*, ``Multiparameter intelligent monitoring in intensive care II (MIMIC-II): A public-access intensive care unit database," *Critical Care Med.*, vol. 39, no. 5, pp. 952_960, 2011.
8. 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.