ISSN (Print) : 0974-6846
ISSN (Online) : 0974-5645
Indian Journal of Science and Technology, *Vol 8(8), 728-730, April 2015*
DOI: 10.17485/ijst/2015/v8i8/69372

Knowledge Abstraction from MIMIC II using Apriori Algorithm for Clinical Decision Support System

E. Nagarajan[1] and V. Satya Sravani[2*]

[1]Faculty of Computing, Sathyabama University, India; enagaraj99@yahoo.com
[2]Department of Computer Science, Sathyabama University, Chennai, India; satyasravani.a@gmail.com

Abstract

Clinical Decision Support Systems provide physicians with a sustainable solution to treat their patients more effectively and improve the quality of care provided to them. These Systems can interpret large volumes of real-time patient data and provide doctors with a snapshot view of actionable information, ultimately allowing them to make better decisions, intervene in a timelier manner. icuARM is one of the Clinical Decision Support Systems, developed so far. This System uses MIMIC II, a publicly accessible Database created to archive records involving multimodal measurement data about ICU patients. icuARM is a tool which uses association rule mining technique that analyzes data for frequent patterns in the database. Overall, icuARM is a CDSS that assists physicians in choosing proper medication based on clinical status of patient's in real time, which will substantially improve the efficiency, accuracy and timeliness for clinical decision making in intensive care. The proposed Apriori association rule mining algorithm generates pattern to associate type of medication to the disease suffered by ICU Patients. The proposed approach decreases the probability of prolonging the patients stay in the ICU and improve their condition. The confidence score achieved in the algorithm is very effective in providing proper medication to the patients.

Keywords: Association Rule Mining (ARM), Clinical Decision Support System (CDSS), Intensive Care Unit (ICU), Multi-Parameter Intelligent Monitoring in Intensive Care II (MIMIC II)

1. Introduction

According to the statistics of SCCM (Society of Critical Care Medicine), there are more than 5 million patients admitted in ICUs annually. Analyzing and integrating this data is the most difficult task for the ICU Physician. CDSS is a trusted system which yields answers clinicians need, at the point of care[1]. Physicians and other health care professionals can use CDSS to prepare and review the diagnosis for critically ill patients. Through CDSS, access to Intensive care information is possible from one place to another. CDSS reduces the risk of medication errors, minimizes the rate of misdiagnosis. It also improves the final outcome by assisting the physician with consistent and reliable information to provide efficient medication for treating ICU patients.

JANANI VENUGOPALAN et al. developed icuARM[1], a Clinical Decision Support System to improve patient safety by the evidence based medicine. It is an "active knowledge system" which uses at least two attributes of patient data records to generate case-specific advice. It can help physicians to examine the symptoms of patients and produce treatment proposal[2]. icuARM uses the knowledge base that contains the rules and associations of data in the form of IF-THEN rules. The rule implies that if a condition exists then the resultant will be notified. The inference engine combines the rules from the knowledge base with the patient's data. The resultant outcome is displayed to the end user by the communication mechanism. The objective of our work is to extract

**Author for correspondence*

rules which reflect the pattern in which medication is given to the existing patients in ICU based on the existing datasets. The decision support system can be used for querying the type of medication to be provided for a new patient by providing his medical profile. The decision support system should provide an insight in improving the prolonged stay in the ICU and recover at the earliest. In children, Antibiotic sensitivity, clinical features are analyzed prior to the medication suggestion and care is taken as they have less resistance power[3].

2. Methods and Procedures

2.1 Association Rule Mining (ARM)

ARM technique was incorporated in areas such as finding patterns in biological databases, survey data from agricultural research, mine data about soil and cultivation, medical databases and so on.

Association rules are of the form A → B, where A is antecedent and B is consequent4. In the ICU data mining process, a rule A → B implies that if A occurs in one ICU stay, B can also occur during the stay. Association rule mining has two basic criteria Support and confidence. These two metrics are slightly modified to fit in accordance with the ICU setting.

Support:

The set AUB defines the occurrences of both sets A and B provided that A and B are not null sets or having missing values. The support can be calculated as follows: (1)

$$supp(A => B) = \frac{count(AUB)}{count((Av \neq \emptyset)U(Bv \neq \emptyset))} \quad (1)$$

Confidence:

The ratio of the items present in both sets A and B to be matched with only the Set A values and set B can contain any values. The confidence can be calculated as follows: (2)

The minimum support threshold and minimum confidence should be specified to find the frequent association rules. At the minimum a rule is frequent if its support and confidence are supp_min and conf_min[4].

$$conf(A => B) = \frac{count(AUB)}{count(A\ U\ (Bv \neq \emptyset))} \quad (2)$$

Algorithm:

```
M = Apriori (S, I, sup)
{
    M1= {m|m∈ I, m.support ≥ sup);
for( K=2;Mk-1≠ Φ ;K++;)
    Cl = Candidate(MK-1);
for each record s ∈ S
    for each candidate c ∈ CK
if (c is present in s)
c.count++;
    Ml={c ∈ CK |c.support ≥ sup}
  return M = UK MK;
}
```

The *Candidate* in the Apriori algorithm is the candidate item set generation algorithm that is given as follows:

```
Candidate (MK-1)
    {
    Cl = {};
    for all mn,mr ∈ MK-1
    {
      mn ={i1,i2,…….il-2,iK-1};
mr ={ i1,i2,……. il-2,i'K-1};
C = { i1,i2,……. il-1,i'K-1};
CK = CK U{c};
for each(l-1)- subset s of c
      if (s ∉ MK-1)
remove c from CK;
    return CK;
        }
```

According to equation (2), if the confidence of a rule is high, it does not necessarily imply that the counter condition of the rule is low confident. When there are two conditions (antecedents) resulting in a single outcome (consequent), then conditions should be taken such that no condition predicts the consequent better than its counter condition. Hence the Importance metric is evolved and can be given as follows: (3).

$$Impo(A => B) = \frac{conf(A => B)}{conf(\bar{A} => B)} \quad (3)$$

Also the rules which cannot predict the consequents better than its counter case are pruned. However, there is a possibility that consequents can be associated with more than one antecedent. In this scenario, the Dominance factor of the antecedent is considered. The metric Dominance can be given as follows: (4)

$$Domi(A \Rightarrow B) = \frac{count(B \cup A)}{count(B \cup (Av \neq \emptyset))} \quad (4)$$

Dominance is similar to (2), the confidence but here the ratio of the items present in both sets A and B to be matched with only the Set B values whatever may beset A can contain any values As there are multiple items present in antecedent of a rule, effect on the confidence of a rule is also determined.

The equation to explain the effect on a confidence of rule after adding new antecedent values can be given as follows:

$$Effe(I_{New} \mid A \Rightarrow B) = conf(A \cup I_{New} \Rightarrow B) - conf(A \Rightarrow B) \quad (5)$$

The range of the effect metric is from -1 to 1. Unlike the support, confidence, importance, and dominance that are all rule-wise metrics, effect is an item-wise metric.

2.3 MIMIC II Database

The data in the icuARM is imported from the MIMIC-II database. MIMIC-II is a research database that can be accessible publicly. MIMIC II database contains records of ICU patients and the medications, diagnosis prescribed to them4. MIMIC-II contains two kinds of data, clinical data and waveform data. Clinical data deals with the patient's symptoms, co morbidities, medications etc. Whereas waveform data mainly focuses on Blood pressure, heart rate, pulse related data etc. In addition to these two types of data, nurse verified charts, fluid balance events was also available[5].

3. Conclusion

icuARM provides real-time data analysis and it is a clinical decision support system which is used to assist physicians in generating qualitative and real-time decision support rules for the ICU setup based on a large ICU patient database MIMICII[6]. The "support" and the "confidence" metrics are modified from conventional form of association rule mining suitable for ICU clinical application. In addition, two new rule-wise metrics "importance" and "dominance" and one item-wise metric "effect" are also developed. An interactive and easy-to-use graphical user interface that enables clinicians to perform flexible data mining in real-time for personalized decision-making[7]. Further, icuARM can be improved as the current mining process with the Apriori algorithm which requires clinicians to manually specify variables of interest and cut-points (for numerical variables) in the items of antecedents and consequents. Instead Automatic feature selection, such as the supervised mRMR method[8] and discretization for more objective item construction can be added.

4. References

1. Cheng C-W, Chanani N, Gopalan JV, Maher K, Wang MD, Cheng et al. icuARM _ An ICU clinical decision support system using association rule mining. IEEE Journal of Transactional Engineering in Health and Medicine. 2013; 2168–372.
2. Thangaraj M, Gnanambal S. A rule based decision support system for aiding vitamin D deficiency management. Indian Journal of Science and Technology. 2014 Jan; 7(1):48–52.
3. Saravanan M, Nanda A. Incidence of Methicillin Resistant Staphylococcus aureus (MRSA) from septicemia suspected children. Journal of Science and Technology. 2009; 2(12):36–9.
4. Agarwal R, Srikant R, et al. Fast algorithms for mining association rules, VLDB '94 Proceedings of the 20th International Conference on Very Large Data Bases.
5. Saeed M, Villarroel M, Reisner AT, Clifford G, Lehman L-W, Moody G, et al. Multi parameter intelligent monitoring in intensive care II (MIMIC-II): a public-access intensive care unit database. Critical Care Med. 2011; 39(5):952–60.
6. Mohod AA, Deshmukh PP, et al. IcuARM-a valuable approach in clinical decision support system. IJAIEM. 2014 Feb; 3(2).
7. Sim I, Gorman P, Greenes RA, Haynes RB, Kaplan B, Lehmann H, et al. Clinical decision support systems for the practice of evidence-based medicine. J Amer Med Inf Assoc. 2001; 8(6):527–34.
8. Peng H, Long F, Ding C, et al. Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundancy. IEEE Trans Pattern Anal Mach Intell. 2005 Aug; 27(8):1226–38.