International Journal of Advance Research in Science and Engineering
Volume No.07, Special Issue No. (01), January 2018
www.ijarse.com
IJARSE
ISSN: 2319-8354

A Modern and Consistent Privacy Protection Scheme For Global Data Based System Using Association Mining Rule in Data Mining Environment

Dr.G.Anandharaj[1], Dr.P.Srimanchari[2], Mr.G.Jayamurugan[3]

[1]Associate Professor and Head, Department of Computer Science,
Adhiparasakthi College of Arts and Science (Autonomous), Kalavai, Vellore (India)
[2]Associate Professor and Head, Department of Computer Applications,
Erode Arts and Science College, Erode(India)
[3]Research Scholar, Part Time External Ph.D,
Bharathiar University, Coimbatore.(India)

ABSTRACT

The prompt enlargement of biomedical monitoring technologies has enabled modern intensive care units (ICUs) to gather vast amounts of multimodal measurement data about their patients. However, processing large volumes of complex data in real-time has become a big challenge. Together with ICU physicians, we have designed and developed an ICU clinical decision support system icuARM based on associate rule mining (ARM), and a publicly available research database MIMIC-II (Multi-parameter Intelligent Monitoring in Intensive Care II) that contains more than 40,000 ICU records for 30,000+ patients. icuARM is constructed with multiple association rules and an easy-to-use graphical user interface (GUI) for care providers to perform real-time data and information mining in the ICU setting. The EU is imposing strict limitations on the use of data obtained from its citizens' online activities [9], while Big Data advocates and online advertisers in the United States are concerned that this may represent interference in their basic business models or even in international trade [13]. It is clear that laws and regulations are inconsistent across national borders. They are also inconsistent within nations, depending on the industry classification of companies, or even the designation given to specific technologies. ISPs are prohibited from reading subscribers' email; other information services companies can do so legally. Data stored electronically is offered protection that is denied to data stored in the cloud. More importantly, it suggests that regulation be driven by what consumers actually want, and provides some preliminary research aimed at determining what consumers want from privacy regulation around the world.

Keywords: Big data, Units, Physician, Cloud computing

International Journal of Advance Research in Science and Engineering
Volume No.07, Special Issue No. (01), January 2018
www.ijarse.com
IJARSE
ISSN: 2319-8354

I.INTRODUCTION

Consumer privacy legislation has received a great deal of attention globally. The EU, Japan, Korea, Malaysia, and many other nations are actively reviewing their policies towards online privacy1. The US has preferred to allow the online information services industry to regulate itself, and the FTC has convened a meeting of the W3C to determine if an acceptable policy could be developed without active involvement by the Congress, the FTC, the FCC, or other regulatory bodies [11,14,17]. To some extent, this activity has been prompted by concerns about what companies like. Our intent with this paper is to provide a sound basis for public policy concerning privacy, in the US and abroad. In order to provide such a sound basis, we conducted surveys and focus groups, in the US, Japan, Korea, and German, concerning consumers' attitudes towards the privacy policies of large information services firms.

We explored both consumers' awareness of specific potential activities in which these large information services firms might engage and consumers' approval of those activities, whether or not firms actually engaged in them. We also explored consumers' attitudes towards the degree of privacy protection they believe they received from their regulatory systems. Finally, we explored parents' attitudes towards data mining of their children's email accounts, including school email accounts. This work as yet has been completed only in the US and Japan.

However, with the advent of the information age, which has led to the explosive growth of information, the scale of graph-based data has increased significantly. For example, in recent decades, with the popularity of the Internet and the promotion of Web 2.0, the number of webpages has undergone rapid growth. Based on statistics provided by the China Internet Network Information Center (CNNIC), at the end of December 2012, the number of webpages in

China had reached 122.7 billion, which represented an increase of approximately 41.7% compared to the previous year. Simultaneously, the number of micro blog users accounted for 54.7% of all Internet users, which is approximately 308 million. This phenomenon highlights the scale of big graph data that is generated, and it is challenging to perform efficient analysis of these data.

To address this issue, we proposed a Bulk Synchronous Parallel (BSP)-based Parallel Big Graph Mining (BPGM) tool. BPGM provides a series of parallel graph mining algorithms based on the BSP method, with the support of distributed file systems for the storage and management of graph data, and a workflow engine to invoke the algorithms.

Our research findings can be summarized as follows:

(1) Consumers mostly do not know what search engines and other data intermediaries are doing with their data, and mostly they know that they do not know.

(2) Consumers mostly do not approve of the majority of the practices of online service vendors when they use consumers' data. This is captured through questions such as, "If it were true that Google did track your searches, would you approve or disapprove?" or "If it were true that your service provider did read your texts, would you approve or disapprove?" The vast majority of consumers disapproved or strongly disapproved of most forms of the use of their online data, no matter what activities resulted in the capture of that data.

(3) Consumers' silence regarding privacy practices and the use of their data by information services firms does

not represent consent, and most certainly does not represent informed consent. We use the term "informed consent" to refer to those consumers who are aware of a practice online and who also approve of that practice. Our data suggests that informed consent is very limited, on the order of 0-1%, for all forms of online privacy abuse. The data are largely consistent across populations surveyed in the US, Japan, Korea, and Germany

(4) Consumers by overwhelming majorities support the position that actually protecting consumers' privacy online should be the default setting on browsers and email services, and on linking online activities with text or GPS information obtained from the user's phone. There was even stronger support for the position that any privacy settings a consumer had chosen to protect privacy, whether set explicitly by the consumer or set implicitly by accepting default settings, should be honored by online service providers. To be clear, consumers believed that the default settings should be do not track individual services and do not integrate across multiple services, and that these default settings should be honored by all online service providers.

(5) Consumers do not believe that regulators are doing enough to inform them about online risks to their privacy and consumers do not believe that regulators are doing enough to protect them from online risks to their privacy.

(6) Consumers have equally strong views about protecting the privacy of their children from data mining activities of information services vendors.

(7) Consumers' feel strongly about protecting the privacy of their children from data mining. They continue to feel strongly even if the email service provider offers email without charge in exchange to the right to perform data mining..

Due to its simple algorithm and good interpretation for recommendations compared to model based methods, similarity based methods have been widely applied, which predict a user's interest for an item based on the weighted combination of ratings of the similar users on the same item or the user on the similar items. The similar users are other users who tend to give similar rating on the same item, while the similar items are the items that tend to get similar rating from the same user.

Since collaborative filtering has been extensively applied in real-world systems, it is meaningful to find other ways to improve its algorithmic performance. Therefore, we propose a Threshold based Similarity Transitivity (TST) method, in which the similarity between two users is not directly computed if their intersection is less than the set threshold and will be replaced by the transitivity similarity.

International Journal of Advance Research in Science and Engineering
Volume No.07, Special Issue No. (01), January 2018
www.ijarse.com
IJARSE
ISSN: 2319-8354

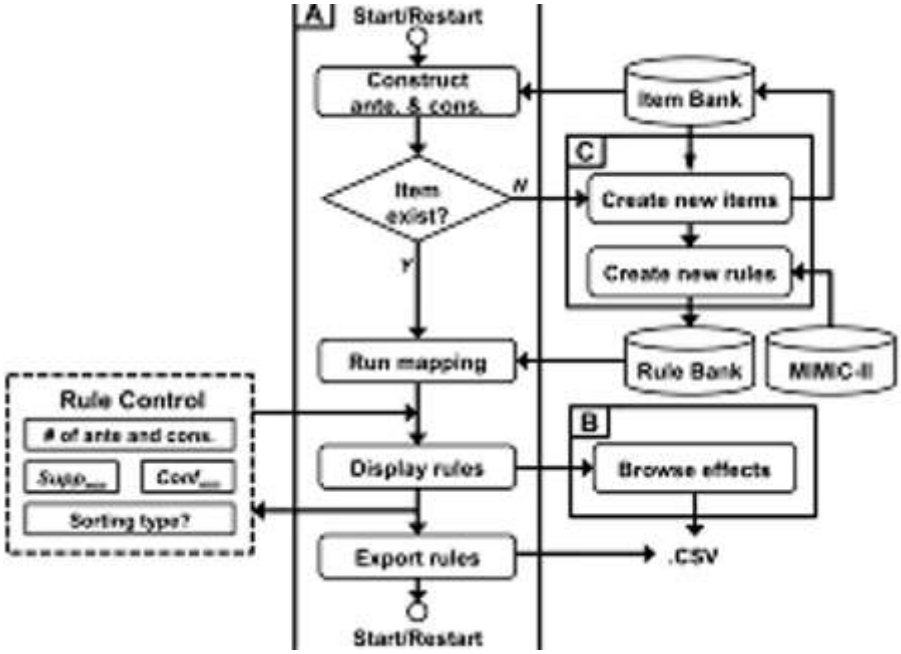


Figure 1.1 Flow of process

II.DEFINING PRIVACY

There are at least three different ways to characterize privacy and online invasions of privacy [2,8,24,38]:

(1) Perhaps the least threatening form of privacy violation is represented by an uninvited intrusion into a user's personal space. Online marketing, spam advertising, pop-ups, and sponsored sites around the edges of a web-page can all be seen as invasions of the user's personal space.

Our focus groups indicated that this was the form of privacy violation most salient in users' minds, across age groups and across nations. When users thought of privacy violations from their email, phone, search, or social network providers, they thought almost exclusively about unwanted ads and unwanted interruptions. That is, they thought of privacy violations in terms of unwanted knocking on a hotel room door when a "Privacy Please" sign was visibly hanging from the doorknob.

(2) Surely the most extreme and most threatening form of privacy violation is represented by fraudulent ecommerce transactions, or even by identity theft. There is no indication that Google, Daum, Navor, Yahoo, or Bing have been associated with such threats to privacy, and there is no indication that consumers were concerned about this.

(3) And surely the form of privacy violations most important to Google, Daum, Navor, and Yahoo are based on personal profiling for some form of commercial advantage. This is definitely an intermediate form of privacy violation. It is far more than simply an invasion of personal space, and far less than identity theft. It involves uniquely identifying an individual and associating him with one or more characteristics of interest to an advertiser. The advertiser's intent may be benign; the firm simply wants to know everyone interested in visiting Osaka, or everyone interested in buying a food processor; this simply results in being sent ads for flights of

interest, or products of interest.

The advertiser's intent may also be less benign; the advertiser wants to know who engages in risky hobbies, so that he can avoid offering life insurance at rates that are too low, or who desperately needs to get to Chicago, so he can offer higher airfares. Interestingly, consumers participating in our focus groups in Japan and Korea initially seemed to focus solely on the first form of privacy violation, the uninvited intrusion into personal space through various forms of spam and targeted marketing. A few hypothetical examples of integration and profiling were sufficient to arouse consumers' concerns.

III.PRINCIPLE OF ASSOCIATION RULE MINING

Association rule learning is a method for discovering interesting relations between variables in large databases. It is intended to identify strong rules discovered in databases using some measures of interestingness.[1] Based on the concept of strong rules, introduced association rules for discovering regularities between products in large-scale transaction data recorded by point-of-sale (POS) systems in supermarkets. For example, the rule $\{\mathrm{onions, potatoes}\} \Rightarrow \{\mathrm{burger}\}$ found in the sales data of a supermarket would indicate that if a customer buys onions and potatoes together, they are likely to also buy hamburger meat. Such information can be used as the basis for decisions about marketing activities such as, e.g., promotional pricing or product placements. In addition to the above example from market basket analysis association rules are employed today in many application areas including Web usage mining, intrusion detection, Continuous production, and bioinformatics. In contrast with sequence mining, association rule learning typically does not consider the order of items either within a transaction or across transactions.

Following the original definition by the problem of association rule mining is defined as:

Let $I = \{i_1, i_2, \ldots, i_n\}$ be a set of n binary attributes called *items*.

Let $D = \{t_1, t_2, \ldots, t_m\}$ be a set of transactions called the *database*.

Each *transaction* in D has a unique transaction ID and contains a subset of the items in I.

A *rule* is defined as an implication of the form: $X \Rightarrow Y$

Where $X, Y \subseteq I$ and $X \cap Y = \emptyset$.

Every rule is composed by two different sets of items, also known as *item sets*, X and Y, where X is called *antecedent* or left-hand-side (LHS) and Y *consequent* or right-hand-side (RHS).

To illustrate the concepts, we use a small example from the supermarket domain. The set of items is $I = \{\mathrm{milk, bread, butter, beer, diapers}\}$ and in the table is shown a small database containing the items, where, in each entry, the value 1 means the presence of the item in the corresponding transaction, and the value 0 represent the absence of an item in a that transaction.

An example rule for the supermarket could be $\{\mathrm{butter, bread}\} \Rightarrow \{\mathrm{milk}\}$ meaning that if butter and bread are bought, customers also buy milk.

Note: this example is extremely small. In practical applications, a rule needs a support of several hundred transactions before it can be considered statistically significant, and data-sets often contain thousands or millions of transactions.

Table 1. Example database with 5 transactions and 5 items

transaction ID	milk	Bread	butter	beer	Diapers
1	1	1	0	0	0
2	0	0	1	0	0
3	0	0	0	1	1
4	1	1	1	0	0
5	0	1	0	0	0

In order to select interesting rules from the set of all possible rules, constraints on various measures of significance and interest are used. The best-known constraints are minimum thresholds on support and confidence.

Let X be an item-set, $X \Rightarrow Y$ an association rule and T a set of transactions of a given database.

1) Support

The support value of X with respect to T is defined as the proportion of transactions in the database which contains the item-set X. In formula: $\mathrm{supp}(X)$

In the example database, the item-set $\{\mathrm{milk}, \mathrm{bread}, \mathrm{butter}\}$ has a support of $1/5 = 0.2$ since it occurs in 20% of all transactions (1 out of 5 transactions). The argument of $\mathrm{supp}()$ is a set of preconditions, and thus becomes more restrictive as it grows (instead of more inclusive).

2) Confidence

The *confidence* value of a rule, $X \Rightarrow Y$, with respect to a set of transactions T, is the proportion of the transactions that contains X which also contains Y.

Confidence is defined as:

$\mathrm{conf}(X \Rightarrow Y) = \mathrm{supp}(X \cup Y)/\mathrm{supp}(X)$.

For example, the rule $\{\mathrm{butter}, \mathrm{bread}\} \Rightarrow \{\mathrm{milk}\}$ has a confidence of $0.2/0.2 = 1.0$ in the database, which means that for 100% of the transactions containing butter and bread the rule is correct (100% of the times a customer buys butter and bread, milk is bought as well).

Note that $supp(X \cup Y)$ means the support of the union of the items in X and Y. This is somewhat confusing since we normally think in terms of probabilities of events and not sets of items. We can

rewrite $supp(X \cup Y)$ as the joint probability $P(E_X \cap E_Y)$, where E_X and E_Y are the events that a transaction contains itemset X or Y, respectively.[3]

Thus confidence can be interpreted as an estimate of the conditional probability $P(E_Y \mid E_X)$, the probability of finding the RHS of the rule in transactions under the condition that these transactions also contain the LHS.[4]

3) Lift

The *lift* of a rule is defined as:

$$\mathrm{lift}(X \Rightarrow Y) = \frac{\mathrm{supp}(X \cup Y)}{\mathrm{supp}(X) \times \mathrm{supp}(Y)}$$

or the ratio of the observed support to that expected if X and Y were independent.

For example, the rule $\{\mathrm{milk}, \mathrm{bread}\} \Rightarrow \{\mathrm{butter}\}$ has a lift of $\frac{0.2}{0.4 \times 0.4} = 1.25$.

4) Conviction

The *conviction* of a rule is defined as $\mathrm{conv}(X \Rightarrow Y) = \frac{1 - \mathrm{supp}(Y)}{1 - \mathrm{conf}(X \Rightarrow Y)}$.

For example, the rule $\{\mathrm{milk}, \mathrm{bread}\} \Rightarrow \{\mathrm{butter}\}$ has a conviction of $\frac{1 - 0.4}{1 - 0.5} = 1.2$, and can be interpreted as the ratio of the expected frequency that X occurs without Y (that is to say, the frequency that the rule makes an incorrect prediction) if X and Y were independent divided by the observed frequency of incorrect predictions. In this example, the conviction value of 1.2 shows that the rule $\{\mathrm{milk}, \mathrm{bread}\} \Rightarrow \{\mathrm{butter}\}$ would be incorrect 20% more often (1.2 times as often) if the association between X and Y was purely random chance.

IV.SYSTEM USE CASES AND INTERFACE

Apriori is an algorithm for frequent item set mining and association rule learning over transactional databases. It proceeds by identifying the frequent individual items in the database and extending them to larger and larger item sets as long as those item sets appear sufficiently often in the database. The frequent item sets determined by Apriori can be used to determine association rules which highlight general trends in the database: this has applications in domains such as market basket analysis.

The GenCandiate in the Apriori algorithm is the candidate

$$
\begin{aligned}
&\text{Apriori}(T, \epsilon)\\
&\quad L_1 \leftarrow \{\text{large } 1-\text{itemsets}\}\\
&\quad k \leftarrow 2\\
&\quad \textbf{while } L_{k-1} \neq \emptyset\\
&\quad\quad C_k \leftarrow \{a \cup \{b\} \mid a \in L_{k-1} \wedge b \notin a\} - \{c \mid \{s \mid s \subseteq c \wedge |s| = k-1\} \nsubseteq L_{k-1}\}\\
&\quad\quad \textbf{for } \text{transactions } t \in T\\
&\quad\quad\quad C_t \leftarrow \{c \mid c \in C_k \wedge c \subseteq t\}\\
&\quad\quad\quad \textbf{for } \text{candidates } c \in C_t\\
&\quad\quad\quad\quad count[c] \leftarrow count[c] + 1\\
&\quad\quad L_k \leftarrow \{c \mid c \in C_k \wedge \ count[c] \geq \epsilon\}\\
&\quad\quad k \leftarrow k + 1\\
&\quad \textbf{return } \bigcup_k L_k
\end{aligned}
$$

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

$$
\begin{aligned}
&\text{Apriori}(T, \epsilon)\\
&\quad L_1 \leftarrow \{\text{large } 1-\text{itemsets}\}\\
&\quad k \leftarrow 2\\
&\quad \textbf{while } L_{k-1} \neq \emptyset\\
&\quad\quad C_k \leftarrow \{a \cup \{b\} \mid a \in L_{k-1} \wedge b \notin a\} - \{c \mid \{s \mid s \subseteq c \wedge |s| = k-1\} \nsubseteq L_{k-1}\}\\
&\quad\quad \textbf{for } \text{transactions } t \in T\\
&\quad\quad\quad C_t \leftarrow \{c \mid c \in C_k \wedge c \subseteq t\}\\
&\quad\quad\quad \textbf{for } \text{candidates } c \in C_t\\
&\quad\quad\quad\quad count[c] \leftarrow count[c] + 1\\
&\quad\quad L_k \leftarrow \{c \mid c \in C_k \wedge \ count[c] \geq \epsilon\}\\
&\quad\quad k \leftarrow k + 1\\
&\quad \textbf{return } \bigcup_k L_k
\end{aligned}
$$

V.RESULTS AND DISCUSSION

5.1 PRE-EXISTING COMORBIDITY VS. PROLONGED ICU STAY

The length of stay (LOS) is a signi_cant ICU outcome that is associated with severe organ failure and high resource consumption [23, 24]. Evidence has shown that patients with prolonged (i.e., longer than 3 days) ICU stays have a considerably increased ICU, hospital, and long-term mortality [25].

Patient comorbidity is a signi_cant variable affecting the ICU LOS [26]. However, survival is typically estimated on a longterm basis (e.g., 1-yr or 2-yrs survival) that is not applicable to the short-term ICU prediction. Therefore, in this case study, we employed the icuARM to generate association rules between pre-existing comorbidities and prolonged ICU stays. comorbidities. The rule with hypothyroidism (HYP) as the comorbidity was ignored because its rule had importance less than 1. Congestive heart failure (CHF) had the highest support (8.6%), which means that this rule was applicable to the highest portion of the ICU stays. Additionally, according to the dominance metric, CHF dominated the prolonged ICU stays by 22.4%, which was also the highest. In this case study, the possibility of prolonged ICU stay can be predicted by the con_dence of a rule given the comorbidities in the antecedent.

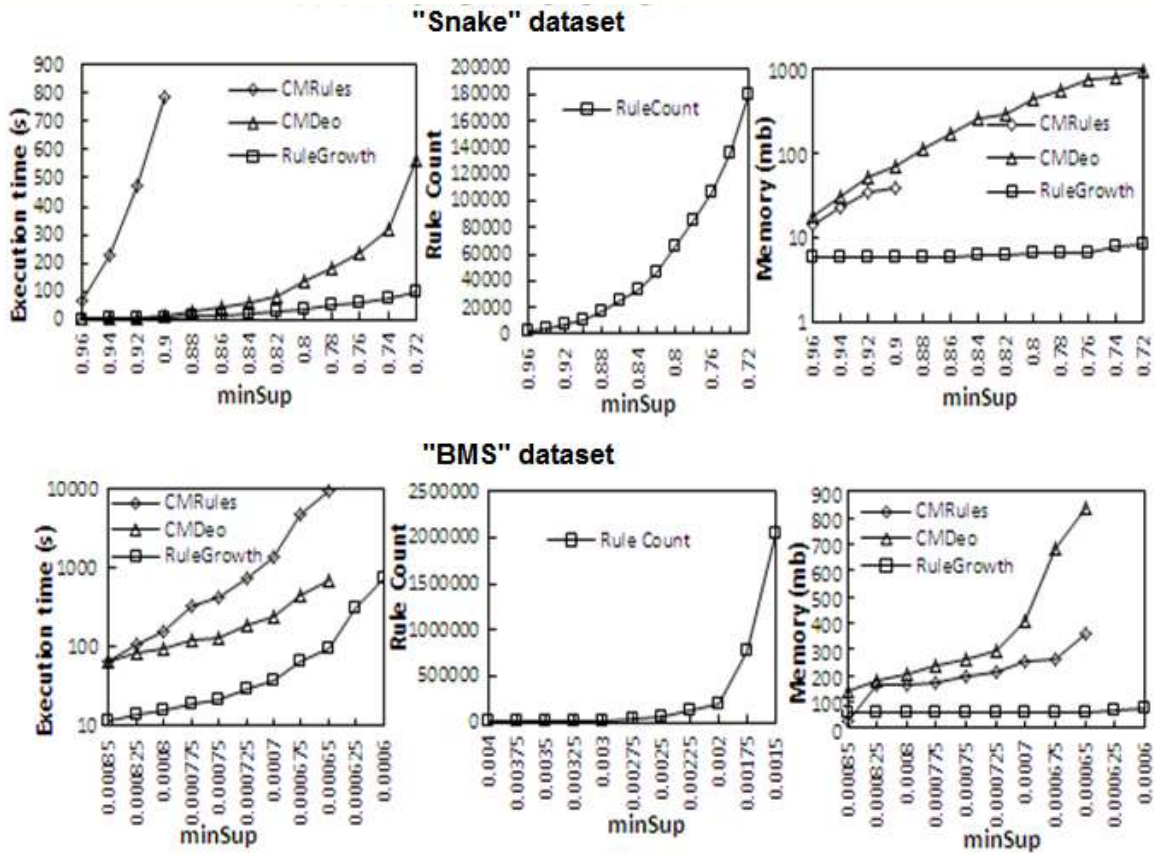
This may be a fact that most of the children within the MIMIC-II database are neonates in the neonatal ICU, where premature infants tend to have prolonged stays (up to months). The icuARM's Effect Browsing window was used to investigate the effect of different combinations of pre-existing comorbidities in different populations on the possibility of prolonged ICU stay. We focused on females aged over 50 because of their highest possibility of prolonged ICU stay. Fig. 5(b) shows all rules of the 11 _rst-item comorbidities in this population. Coagulopathy (COA) was still associated with the highest possibility (58.6%) of prolonged ICU stay. In addition, females over 50 years who had alcohol and/or drug abuse showed an increased possibility (44.1%) even though this comorbidity did not have a high risk in the general population.

We continued to investigate the effects (i.e., changes of possibility of prolonged ICU stay) of possible second-item comorbidities in females aged over 50 years who also had coagulopathy. there were 10 possible second-item comorbidities that were important (importance

1). Among them, six comorbidities increased the possibility, with de_ciency anemia (DEA) resulting in the highest rate of prolonged ICU stay (64.4%). Clinicians can continue the effect browsing process by adding other comorbidity combinations based on a patient's status at admission.

Some comorbidities tend to be associated with shorter ICU stays, especially in younger population. For example, young patients with ABU usually admit for short (i.e., < 24 hrs) ICU stays because of acute alcohol intoxication (e.g., seizures/delirium tremens) or drug overdose/intoxication (e.g., respiratory monitoring after an opiate or benzodiazepine overdose).

International Journal of Advance Research in Science and Engineering
Volume No.07, Special Issue No. (01), January 2018
www.ijarse.com
IJARSE
ISSN: 2319-8354



Possibility of prolonged ICU stays in different age-gender populations. The three values of each bar are the three measures of the association rules, including support (top), importance (middle), and dominance (bottom). The effects (i.e., changes in possibility of prolonged ICU stay) of the first-item comorbidities on prolonged ICU stay possibility in females aged over 50 years. The effects of the second-item comorbidities on prolonged ICU stay possibility in females aged over 50 years who also have coagulopathy. After evaluating rules based on their support and condense values, it is important to interpret the importance value.

Rules have lower importance values for more general cases with fewer items (in either the antecedent or consequent), and have higher importance values for more cases with more items. When comparing the importance values among

We can observe that the importance values increase as more items are added to the antecedents. We want to emphasize, so even if, theoretically, the value can go to in_nity, in reality. In this case study, we have shown the basic usability of icuARM to assess associations between pre-existing comorbidities and prolonged ICU stays, especially in females aged over 50 years. Clinicians can construct different combinations of age, gender, and pre-existing comorbidities to determine a baseline prolonged ICU stay possibility of a patient at the time of ICU admission, even prior to diagnosis. By estimating the possibility of prolonged ICU stay, an ICU team can plan ahead for the intensive care resource allocation such as laboratory, and radiology. This prediction also provides a risk reference to assess how certain interventions will affect LOS. For example, a clinician may admit two female patients of similar age. One has a coagulopathy (e.g., disseminated intravascular coagulation) and the other has an acute coronary syndrome. By using icuARM, the clinician and ICU team could accurately plan for

needed resources for the former patient, predict outcomes, and improve management for this type of high risk ICU patient.

VI. MEDICATION USAGE VS. PROLONGED ICU STAY

Mining associations between medication usage and clinical outcome is another promising application of icuARM. ARM has been adopted in several studies, such as investigating multi-item adverse drug reactions [27-29]. However, to our knowledge, no CDSSs have adopted ARM for associations between medication usage and ICU outcomes. Therefore, in our second case study, by using icuARM, we investigated the associations between prolonged ICU stays and medication usage in addition to patient demographics and pre-existing comorbidities. We are mined the association rules of two commonly used anti-hypertensive drugs in ICUs: diltiazem (DIL) and labetalol (LAB). We selected males and females over 50 years because they had the highest prolonged ICU possibility according to our previous case study. The associations on the drugs with a pre-existing comorbidity of congestive heart failure (CHF) were also investigated. In patients over 50 years without CHF, DIL is associated with higher possibility compared to LAB in both females and males. However, these two drugs had different effects on patients with CHF. For females over 50 years with CHF, the use of DIL increased the possibility of prolonged ICU stays to 83.4% compared to females over 50 without CHF (73.6%), whereas LAB had nearly no change (62.3% vs. 61.7%). In contrast, for the same clinical situation, the use of LAB actually increased the possibility in males over 50 years with CHF to 87.1% compared to those without CHF (62.8%), whereas DIL had almost no effect (74.7% vs. 76.0%). Therefore, for patients over 50 years with a comorbidity of CHF, we may choose LAB for females and DIL for males. Medication usage Vs. prolonged ICU stay (>50 years old).

In addition to the hypertensive conditions, ICU clinicians often have a choice between pharmacologic agents in an acute episode of cardiopulmonary arrest. Epinephrine (EPI) and vasopressin (VAS) are two common drugs used in the management of ventricular and pulseless electrical activity. In addition, Gueugniaud *et al.* suggested that the combination of EPI and VAS did not improve outcome (i.e., survival to hospital discharge, good neurologic recovery, and 1-year survival) during advanced cardiac life support for out-of-hospital cardiac arrest [30]. However, evidence was still to make prognosis on short-term ICU stays. Therefore, by utilizing icuARM, the association between ICU LOS and a combination of EPI and VAS was also evaluated compared to EPI or VAS alone.

According to the result shown in females over 50 years without CHF had slightly lower possibility of prolonged ICU stay with VAS compared to EPI (64.7% vs. 67.6%); in contrast, males over 50 without CHF had lower chance of prolonged ICU stay with EPI compared to VAS (61.8% vs. 71.4%). These associations all increased on patients over 50 who also had CHF, but the EPI increased the possibility most on females (84.5%) compared to those without CHF (67.6%).

Furthermore, for the combination of EPI and VAS, the change of the possibility was not considerably different compared to EPI or VAS alone. This partially supported the _nding of [30] although we focused on the short term ICU outcome. This case study demonstrated that icuARM could help guide the clinician to select correct

medication for similar clinical situations but different patient populations in the preplanning phase. The entire mining process requires no more than one minute, promising a nearly real-time and easy-to access bedside consulting tool.

VII.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 , there is a critical need to apply modern information technology and advanced data analytics to extract information from heterogeneous clinical data.

We adopted the ``support" and the ' metrics suitable for ICU clinical application from conventional association rule mining. In addition, we de_ned and developed two new rule-wise metrics ``importance" and ``dominance" and one item-wise metric ``effect." We developed an interactive and easy-to-use graphical user interface that enables clinicians to perform _exible data mining in real-time for personalized decision-making. We tested icuARM on two cases investigating the associations between prolonged ICU stays and patient demo graphics ,pre-existing comorbidities, and medication usage. Our results not only reinforced the current decision-making evidence, but also revealed new knowledge by predicting characteristics of a prolonged ICU stay. We will further improve this CDSS in four directions. First, besides basic patient demographics, pre-existing comorbidity data, and medication usage, we will emulate more categories such as nurse-veri_ed chart events, laboratory tests, and _uid balance records etc. to better assist ICU clinicians in making critical decisions. Second, as mentioned in we aim to include continuous physiological data with corresponding time stamps from the MIMIC-II database to perform temporal association rule mining [31].

Third, 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 [32] or the unsupervised MCFS method [33], and discretization for more objective item construction. Finally, in addition to the _ve evaluation metrics, we will provide a more comprehensive clinical evaluation by including other data centric rule metrics [34].

Consumers' preferences would suggest that the default privacy settings should be no tracking without explicit permission and no integration without explicit permission. Governments should hold firms accountable for violations and should ensure that violations are visible when they occur. Governments should ensure that consumers can know exactly what companies have done, so that they can protect themselves, register displeasure, and change vendors if they deem it necessary to stop violations of their privacy. Our data also show that the public is not widely aware of criminal violations of privacy, such as the WiSpy scandal in the US or iPhone hacking in the US, or Google Analytics violation of EU privacy laws. Given that these violations are strongly counter to consumers' preferences, perhaps they should receive greater attention and be more severely punished when they occur.

REFERENCES

[1] A. Rauber, ``LabelSOM: On the labeling of self-organizing maps," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, vol. 5. Jul. 1999

[2] 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.

[3] 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.

[4] 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.

[5] D. L. Sackett, ``Evidence-based medicine," *Seminars Perinatol.*, vol. 21, no. 1, pp. 3_5, 1997.

[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] W. A. Knaus, E. A. Draper, D. P.Wagner, and J. E. Zimmerman, ``APACHE II: A severity of disease classi_cation system," *Critical care Med.*, vol. 13, no. 10, pp. 818_829, 1985.

[8] M. Verduijn, N. Peek, F. Voorbraak, E. De Jonge, and B. de Mol, ``Dichotomization of ICU length of stay based on model calibration," in *Arti_cial Intelligence in Medicine*. New York, NY, USA: Springer-Verlag, 2005, pp. 67_76.

[9] F. L. Ferreira, D. P. Bota, A. Bross, C. Mélot, and J.-L. Vincent, ``Serial evaluation of the SOFA score to predict outcome in critically ill patients," *J. Amer. Med. Assoc.*, vol. 286, no. 14, pp. 1754_1758, 2001.

[10] G. Teasdale and B. Jennett, ``Assessment of coma and impaired consciousness: A practical scale," *Lancet*, vol. 304, pp. 81_84, Jul. 1974.

[11] K. A. Fox, O. H. Dabbous, R. J. Goldberg, K. S. Pieper, K. A. Eagle, F. Van de Werf, *et al.*, ``Prediction of risk of death and myocardial infarction in the six months after presentation with acute coronary syndrome: Prospective multinational observational study (GRACE)," *BMJ*, vol. 333, no. 7578, pp. 1091_1906, 2006.

[12] E. Turban, J. Aronson, and T.-P. Liang, *Decision Support Systems and Intelligent Systems*, 7th ed. Englewood Cliffs, NJ, USA: Prentice-Hall, 2005.

[13] J. Ramon, D. Fierens, F. Güiza, G. Meyfroidt, H. Blockeel, M. Bruynooghe, *et al.*, ``Mining data from intensive care patients," *Adv. Eng. Informat.*, vol. 21, no. 3, pp. 243_256, 2007.

[14 J. Dean and G. Sanjay, MapReduce: Simplified data processing on large clusters, Communications of the ACM, vol. 51, no. 1, pp. 107-113, 2008.

[15 S. Marc, S. W. Otto, D. W. Walker, J. Dongarra, and S. Huss-Lederman, MPI: The Complete Reference. MIT press, 1995.

[16] L. G. Valiant, A bridging model for parallel computation, Communications of the ACM, vol. 33, no. 8, pp. 03-111, 1990.

[17] Acquisti, A., John, L. and Lowenstein, G. 2009. What is privacy worth? Workshop on Information Systems and Economics (WISE).

[18] Altman, I. The Environment and Social Behavior: Privacy, Personal Space, Territory and Crowding. Brooks/Cole Pub. Co., Inc., Monterey, CA, 1975.

[19] Angwin, Julia. (2010, July 10). "Google, FTC Near Settlement on Privacy." The Wall Street Journal. http://online.wsj.com/article/SB10001424052702303567704 577517081178553046.html.

[20] Bowdon, Bob. (2011, Feb. 22). "Why has Google been Collecting Kids' Social Security Numbers under the Guise of an Art Contest?" Huffington Post. http://www.huffingtonpost.com/bob-bowdon/why-hasgoogle- been-colle_b_825754.html.

[21] Clemons, E., JIN F., Wilson, J., REN, F., Matt, C., Hess, T., and KOH, N. "The Role of Trust in Successful Ecommerce Websites in China: Field Observations and Experimental Studies." Proceedings, 45th International Conference on System Sciences, Maui, Hawaii, January 2013.

[22] Department of Justice. (2011, Aug. 11). Google Forfeits $500 Million Generated by Online Ads & rescription Drug Sales by Canadian Online Pharmacies. Press Release. http://www.justice.gov/opa/pr/2011/August/11-dag-1078.html.

[23] Electronic Privacy Information Center. "Student Privacy." http://epic.org/privacy/student/.

[24] Etzioni, A. The Limits of Privacy. Basic Books, New York, 1999.

[25] European Commission. (2013, Feb. 20).MEMO/13/124, "EU Data Protection: European Parliament's Industry committee backs uniform data protection rules." Press Release. http://europa.eu/rapid/press-release_MEMO- 13-124_en.htm.

[26] Family Educational Rights and Privacy Act. 20 U.S.C. § 1232g; 34 CFR Part 99.

[27] Federal Trade Commission. 2009. Self-Regulatory Principles for Online Behavioral Advertising. http://www.ftc.gov/os/2009/02/P085400behavadreport.pdf.

[28] Federal Trading Commission. 2011. "FTC Charges Deceptive Privacy Practices in Go

[29] National Research Council, Frontiers in Massive Data Analysis. Washington, DC, USA: National Academies Press, 2013.

[30] J. E. Kelly and S. Hamm, Smart Machines: IBM's Watson and the Era of Cognitive Computing. New York, NY, USA: Columbia Univ. Press, 2013.

[31] J. G. Wolff, ``The SP theory of intelligence: An overview," Information, vol. 4, no. 3, pp. 283 341, 2013.

[32] J. G. Wolff, Unifying Computing and Cognition: The SP Theory and Its Applications. Menai Bridge, U.K.: CognitionResearch.org, 2006.

[33] J. G. Wolff, ``The SP theory of intelligence: Bene ts and applications,' Information, vol. 5, no. 1, pp. 1 27, 2014.

[34] M. Li and P. Vitányi, An Introduction to Kolmogorov Complexity and Its Applications. New York, NY, USA: Springer-Verlag, 2009.

[35] J. G.Wolff, ``Information compression, intelligence, computing, and mathematics," 2013, in preparation.

[36 J. G.Wolff, ``The SP theory of intelligence: An introduction," unpublished, Dec. 2013.

[37] T. Berners-Lee, J. Hendler, and O. Lassila, ``The semantic web," Sci. Amer., vol. 284, no. 5, pp. 35 43, May 2001.

[38] P. Bach-y-Rita, ``Theoretical basis for brain plasticity after a TBI," Brain Injury, vol. 17, no. 8, pp. 643 651, 2003.

[39] P. Bach-y-Rita and S. W. Kercel, ``Sensory substitution and the humanmachin interface," Trends Cognit. Sci., vol. 7, no. 12, pp. 541 546, 2003.

[40] J. G. Wolff, ``Towards an intelligent database system founded on the SP theory of computing and cognition," Data Knowl. Eng., vol. 60, no. 3, pp. 596 624, 2007.

[41] J. G. Wolff, ``Medical diagnosis as pattern recognition in a framework of information compression by multiple alignment, uni cation and search," Decision Support Syst., vol. 42, no. 2, pp. 608 625, 2006.