Ontology Based Context-aware Post-mining of Association Rules

R. Subashchandrabose¹ and R. Sivakumar²

¹Research Scholar, Department of Computer Science, Bharathiyar University, Coimbatore, Tamil Nadu, India.
²Associate Professor and Head, Department of Computer Science, AVVM Sri Pushpam College, Poondi, Tamil Nadu, India.

Abstract
Data mining has been widely accepted as a tool for effective decision-making. In data mining, the usefulness of association rules is limited by the volume of generated rules based on statistical measures. Ontology is considered as the most appropriate representation to express user domain knowledge to guide the association rule mining process in multi-database at multilevel at the post-processing step to reduce the number of rules to a few. In human-computer interaction, inputs can be captured via explicit models of communication. However, implicit context factors play a role to enrich user inputs and thereby improve the performance of post-mining of association rules mining further. We propose a Context-aware Rule Schema multi-Level formalism for user expectations and context factors. Furthermore, this paper proposes Context-aware post-mining of association rules framework by which context will be considered and that produces more accurate and precise rules. By applying our new approach over educational domain. We observed that our new framework was adapted to various context factors and suggested implications pertinent to the nature of context.

Keyword: Context-aware, multilevel, rule schema, ontology, post-mining, user knowledge.

INTRODUCTION
Data mining is the process of extracting interesting pattern or knowledge from a database of large collections of past historical data and it has been widely used as a tool to assist the decision-maker. Because of the very fast development in the field of information and communication technologies, data may be distributed over geographically different locations. In such situations, the multi-database mining plays an important role in the process of extracting knowledge from different data sources of any domain of interest with the help of local mining techniques. Afterwards, the discovered knowledge can be used locally and forwarded to the central knowledge base, where patterns or knowledge undergo an evaluation, called synthesized process to rank them according to the usefulness of the pattern or knowledge. A pattern can be a frequent item set or an association rule. In the literature, we found so many interesting research papers on multi-database association rule mining have been presented in [1] [2] [3] [4] [5] [6] at the post-mining stage. Further, mining the multi-database in the presence of context factors may improve the performance of mining association rule. This paper proposes a new framework, in which context will be considered to maximize the adaptive capability of the Knowledge Discovery in Databases (KDD) process. The context can be of any circumstantial factors of the user and domain that may affect the mining process. Context can be used to interpret and enrich user inputs and thereby affecting association rule mining results. We focus on building a context-aware post-mining of association rule framework in multi-databases at multilevel that can filter useful and interesting context factors, which produce accurate and precise prediction results. The framework was proposed in keeping the educational domain in mind. This paper is organized as follows. Section II discusses the motivations for using ontologies and context-aware rule schema. Section III introduces the related work in ontology-based association rule mining process. Section IV describes the detail of the proposed framework and its components. A case study and its implications are presented in Section V over educational domain. Finally, the last Section presents conclusions and shows direction for future research.

MOTIVATIONS FOR USING ONTOLOGIES AND CONTEXT-AWARE RULE SCHEMA IN POST-MINING PROCESS
Post-mining of Association Rules
The post-mining process, which aims to reduce the number of rules with an efficient post-processing methods, is a crucial step in KDD process to help the decision-maker with a few rules. Several algorithms were introduced in the literature to reduce the number of item sets and several algorithms to reduce the number of association rules by pruning techniques or using no redundant rules. On the other hand, post-processing methods can improve the selection of interesting rules. However, most of the post-processing methods are generally based on statistical information in the database. Those methods do not give any guarantee that interesting rules will be selected. User-driven post processing methods which focus on specific schemas of rules to select such interesting ones. The representation of user knowledge is an important issue to represent the user expectations in a flexible, expressive and accurate formalism by which rule selection process becomes more efficient. Thus, ontology is the most suitable representation of the user knowledge.
Ontologies in data mining

The term ontology originates from philosophy that deals with existence of things that exist. In Knowledge Engineering, things that exist are those which can be represented by data. Ontology can be defined as a formally defined system of concepts and relation between these concepts. In 1990s, Gruber [22] defined ontology as a formal, explicit specification of shared conceptualization. Here, conceptualization refers to an abstract model of some phenomenon in the world described by its important concepts or terms. The formal notion denotes the fact that ontology is machine readable. Finally, shared outlines that an ontology brings together some knowledge common to certain group and not a private of any individual.

Domain knowledge is defined as the user background knowledge concerning the database that is modeled which relies on using ontology.

General Impressions, Reasonably Precise Concepts and Precise Knowledge were derived from the specification language proposed by Liu. B. et. al. [21]. Taxonomies used only Is-a relation that is extended with the set of relations R in order to have a more complex knowledge representation. In addition, axioms are very useful to infer new knowledge by permitting some important improvements on the definition of concepts.

Context-Aware

According to Dey and Abord, context is defined as a piece of information that can be used to characterize the environment of participants in an interaction. Similarly, context is a as information on the location, time, environment, identity of people in a certain situation. The concept of context has been interwoven with many different applications. An information may be interpreted as meaningful in one context and the same information as meaningless in another context. Thus, when the information has to be conveyed from one element to another, the interpretation of information can be changed particularly depending on the context of information. Context enabled applications can modify their behavior according to changes in the environment by sensing context information. Applications can use different types of context factors: user context, application context, time context, data context, data mining context, attributes context and performance context.

Pravin Vajirkaret al. proposed a Context-Aware Data Mining Framework [23], which considered context factors in the mining process. They proposed context-aware process component, ontology component, data mining component and user interface component in their study. Context-aware component handles context factors for data mining and query processor components. In this component, they considered the context such as application text, user context, domain context, data context. Based on the context-aware factors, discovered rules are classified with minimum user’s involvement.

RELATED WORK

Data mining is used to discover hidden knowledge or pattern in large databases. Traditional association rule mining is based on objective measures which extracts frequently occurring item set in the form of association rules which is not considered the meaning of each data item. On the other hand, subjective measures require user involvement in finding association rules. Ontology play important roles at various levels of KDD process and it is used as user knowledge representation about a domain. The semantic content extracted from the ontologies allows inserting more intelligence and knowledge in data mining process and improving their quality.

Ontologies were introduced in data mining in early 2000 that can be used in different ways[7]: Domain ontologies, Application ontologies, Background ontologies, ontologies for Data mining process and Metadata ontologies. The activities of building ontologies are identifying relevant concepts, building is-a hierarchies and defining relationships among concepts. The main task of ontologies for data mining process is to codify mining process description and choose most suitable task according to the nature of the given problem. Metadata ontologies are used to describe the construction of items[8].

In this survey, we are interested to present the past studies related to domain knowledge ontologies. Domain ontology was introduced by Srikanth and Agarwal with the concept of generalized association rules[9]. Cumulative and Est Merge algorithms were proposed by the authors to find association between items at any level of hierarchy. In this research, candidates item set are formed by adding items of different level.

Cespivova et al. introduced[10] medical ontology and other background knowledge into the process of association rule mining. The authors used LISP-Miner tool, the UMLS ontology, the STULONG dataset on cardiovascular risk and a set of simple quality association rules for their experiment. The observation from the experiment was that ontology may bring the benefits to all the phases of Knowledge Discovery in Databases (KDD) process cycle as followed in CRISP-DM. Euler. T and Scholz. M presented in [11] a metamodel of KDD preprocessing steps that uses ontology is operational, yet abstract enough to allow the reuse of applications in similar domains. They proposed MiningMart system which provides support for modeling conceptual knowledge about a domain for finding the right representation of data for the mining process. In the MiningMart system, domain ontology structures the relevant concepts of a domain by means of inheritance relationship and other kinds of relationships between concepts. This kind of domain model allows for convenient handling of views in different contexts. This will improve the interpretability and reusability of KDD processes.

Charest. M and Delisle developed [12] ontology guided method for data mining using case-based reasoning by the realization of a hybrid intelligent data mining assistant based on the combination of both declarative and procedural ontology knowledge. It improves the power of the non-specialist data miner throughout the key phases of the CRISP-DM data mining process.

Zohu. X and Getler. J developed [13] a Raising method that used to perform KDD pre-process steps on a dataset before applying any data mining algorithm. The Raising method exploits the taxonomy of the ontology and adds instances at the lower levels of taxonomy to enhance the association rule derivation process which involves the ancestors of these
instances. Raising is the process of generalizing association rules in order to increase the threshold value support while it keeps threshold confidence value enough high so that it guarantees to generate better rules. The difference between Raising concept and generalized association rules concept is that Raising allows the user to use a specific level for raising. The concept fuzzy set theory [14] applied a lot on data mining especially on mining association rules in the literature. Mining generalized association rules based on fuzzy ontology approaches result semantically richer information with a certain amount of redundant rules. We also found many association rule mining algorithms based on fuzzy ontology. The algorithm Semantically Similar Data Miner (SSDM) [15] considered not only exact matches between items and it also considered semantic similarity between items. Escovar. E et. al. extended [16] the SSDM algorithm in order to establish semantic relations between items from a fuzzy ontology. SSDM uses fuzzy logic concepts to show the similarity degree between items and proposes a new way of obtaining threshold values. Generalized association rule mining methods based on fuzzy ontology produce semantically richer information despite it produces abundant redundant rules. Thus, treating redundant rules is another interesting area of research. The Extended SSDM (XSSDM) can reuse consensual and shared knowledge to simplify the process of acquiring semantic contents. Mining Non-redundant Association Rules based on Fuzzy Ontology (NARFO) algorithm [17] proposed to obtain non-redundant association rules. NARFO algorithm extracts important meaningful knowledge from non-frequent itemsets. It also contributes at the stage of post-processing with its capabilities of generalization and redundancy treatment.

Mansingh. Get. al. presented [18] a hybrid method for processing association rules in order to reduce the cognitive burden on the knowledge workers. The method combined the both domain knowledge and objective measures to extract and partition interesting patterns and knowledge from databases into four categories: known, novel, contradictory and missing association rules. The authors demonstrated the method in a medical domain data set and showed that it provides a mechanism for reusing and automatically updating knowledge base.

Manda. P et. al. proposed a strategy called Multi-Ontology and All Level (MOAL) algorithm[19]. They make use of the structure and relationships of a Genetic ontology to mine multi-ontology multilevel association rules. Multi-ontology support and Multi-ontology confidence are the two measures used to evaluate multi-ontology multilevel association rules. It also offers a wide variety of strategies for pruning uninteresting rules.

Neves and Ana presented [20] a study on the effect of using a preparation technique, which is defined in terms of precision and recall, called SemPrune is, based on domain ontology. SemPrune technique is used at the stages of pre-processing and post-processing of KDD process. The main aim of the study is to identify generalization/specialization relations as well as composition/decomposition relations that is a key to applying SemPrunetechnique successfully.

Liu. Bet. al. proposed [21] a framework, which has two components: interestingness analysis component and visualization component, allows the user to identify the interesting ones by exploring the discovered association rules. The discovered association rules are analyzed by the interesting analysis component according to various interestingness measures with respect to users knowledge. The visualization component allows the user to explore potentially interesting rules visually.

Claudia Marinica and Fabrice Guillet proposed a novel framework [8]Association Rule Interactive and Post-processing using Schemas and Ontology (ARIPSO) to analyze the rule in an interactive manner to prune and filter the discovered rules. They addressed the two main issues: first, the integration of user knowledge based on ontology that is used at the stage of post-processing, second, the interactive framework to assist the user for which they propose rule schema formalism throughout the analyzing task and a set of filtering operators to compare the discovered rules with rule schemas. The framework was tested successfully over the client database provided by Nantes Habitat.

Salim Khiat et. al. proposed a framework [24]Multilevel abstraction of Association rule using Ontology and Rule schema (MAROR) to mine association rules in multi-database for a multilevel and multi-nation organization of a company distributed over different geographical locations. In the framework, a rule-like formalism called Rule Schema Multi-Level, was introduced to specify user expectations that acts as rule grouping or defining family of rule. First, they introduced ontology which allows the user to express his/her domain knowledge by means of high semantic model. Second, they introduced a Rule Schema Multi-Level filtering model which allows the user to define his/her expectations regarding the rules. The framework consists of Extern-site and Auto-site processes that guide the mining process at the stages of post-processing and data mining. Rules are synthesized based on user knowledge into majority, global and exceptional. These synthesized rules for a multilevel organization are helpful to the decision-maker of a company. It is demonstrated in the maintenance process of a petroleum company for detecting equipments failure which will reduce the cost of maintenance. In this paper, we propose a new framework, which can be applied to mine association rule at the post-processing step in multi-database. We propose a context-aware rule-like formalism, called Context-aware Rule schema, to represent user knowledge or belief and context-factors. This schema allows the user at different level of expertise in their domain of interest to define their own expectations regarding the rules by means of ontology concepts. We propose a new set of operators over each context-aware rule schema for interactive and iterative processing. We also propose a new synthesizing process to generate semantically valid rules with respect to user knowledge. Finally, the framework was tested in educational domain to provide the more useful content without any redundant and irrelevant content of information depends upon context-awareness level.

**CONTEXT-AWARE POST-MINING OF ASSOCIATION RULES FRAMEWORK**

In this paper, a new framework (CPAROR) (Context-aware Post-mining of Association Rules using Rule schema) is
proposed to select the interesting association rules by the knowledge workers at different level of abstraction. CPAROR works locally as well as globally in multi-database association rule mining process. We propose the framework for multi-database multilevel organization to select mostly the interesting rules in two phases of mining process. First, in auto-site phase, association rules are generated by integrating user knowledge and context factors in the mining process. Second, extern-site phase mines some interesting association rules for each level of the multilevel organization. The proposed CPAROR framework is presented in figure 1.

A semantic model is introduced in CPAROR, called ontology, to represent user domain knowledge. We propose a new rule-like formalism in this framework called Context-aware Rule Schema Multi-Level, which combines user-knowledge and context factors, thereby improving the performance of post-mining of association rule mining process. In other words, lack of context-awareness would affect this mining process. Finally, applying a new set of operators applied over each schema interactively to select accurate and more precise prediction rules.

Ontology Definition 1
In this framework, it is fundamental to connect ontology concepts C of $O = \{C, R, I, O, H\}$ to the database. For this purpose, we consider three types of concepts: leaf-concepts ($C_0$), generalized concepts ($C_1$) from the subsumption relation ($\subseteq$) and restriction concepts ($C_2$) defined by ontology. The concept $C$ of $O$ is defined as union of three subset of concepts as $C=C_0 \cup C_1 \cup C_2$.

The subset $(C_0)$ is defined as the set of leaf-concepts of the ontology connected to the database. $C_0=\{c_0 \in C | \exists c' \in C, c' \subseteq c_0\}$

In this manner, each concept of $C_0$ is connected to an item in the database.

Generalized concepts ($C_1$) are described as the concepts subsume other concepts of the ontology. A generalized concept ($C_1$) is connected to the database through subsumption relation. This means that only the leaf-concepts subsumed recursively by the generalized concepts contribute to its database connection.

$\forall c \in C_1, f(c)=\{i= f_0(c_0) | c_0 \in C_0, c \subseteq c_0\}$

Restriction concepts ($C_2$) are described by using logical expressions defined over items which are organized in $C_2$ subset. In a first attempt, we have the description of concepts on restrictions over properties available in logics. Thus, restriction concepts could be connected to a disjunction of item.

$\forall c \in C_2, c \rightarrow E(c)$

Definition 2
A Context aware rule schema is defined as:

$\langle \text{rdfs: Literals} \rangle \wedge \langle X_1, X_2, ... X_m \rightarrow (Y_1, Y_2, ... Y_k) \rangle$ (T)(N)

Where:

- $\wedge$ refers to a conjunction operators to wrap up context factors and rule schema
- $X_i$ and $Y_j$ are ontology concepts and the implication $\rightarrow$ is optional.
- T = {L, M, E, G} is the type of knowledge which can be local(L), majority(M), exceptional(E) and global rules(G).
- N is the level of the context aware rule schema which indicates the level of users that formulate this one. The lower level (n=1) exposes the decision maker’s belief in the lower organization level. The higher level $n$ (n>1) expresses the decision maker’s belief in the middle and head quarter of the organization.

If the implication $\rightarrow$ ‘ is mentioned in the context aware rule schema, we say that this schema is an implicative for reasonably precise concepts and precise concepts and if we do
not use the implication $\rightarrow^*$, it is called non-implicative for
general impressions.
For example, a context-aware rule schema,
$<\text{user\_context}(x),\text{location\_context}>^<\left(C_1, C_2 \rightarrow C_3\right)>^\text{(M)}$

Correspond to all Majority type Association Rules with
respect to the user and location contexts and the ontology
concepts $C_1$, $C_2$ and $C_3$ at level 2.

Operators
Our aim is to design a post-processing task that is based on
the operations applied over context-aware rule schema which
is allowed to select the interesting rules by the user. For this
purpose, we propose two kinds of operators: Auto-site and
Extern-site.

Definition 3
Let us consider an ontology concept $C$ associated in the
database to $F(C) = \{y_1, \ldots, y_n\}$ Where $\{y_1, \ldots, y_n\} \in I$ and an
itemset $X = \{C_a\} \cap \{x_1, \ldots, x_m\}$. We say that the itemset $X$ is
conforming to the concept $C$, if $\text{conf}(X, C) = \text{TRUE}$, if $\exists y$
and if $\text{conf}(X, C) = \text{FALSE}$, otherwise

Extern-Site Step
The process of $\text{CPARORExtern-site}$ framework is shown in
figure 3 is to guide the user at the post-processing stage. The steps are suggested as follows:

![Figure 3: Extern-site Process](image)

1. Ontology construction: The first step is to map the
given database into ontology concepts.
2. Context-aware Rule Schema definition: Transform
the ontology concepts into context-aware rule
schema based on the user knowledge and context-
aware concerning the mined association rules.
3. Operators’ definition: Choosing the operator is
strongly depends on the user privilege which will be
helpful to choose the actions to be applied over the
context aware rule schema in this regard, we
introduce context factors in the rule schema
definition of $\text{MAROR}$ framework[24] that is adapted
in our approach.
4. Rule Synthesizing: The rule synthesizing process
applied to generate meaningful rules with respect to
user knowledge and context factors. Rules are
clustered into Global(G), Majority(M) and
Exceptional(E) which are more useful for a multi-
level organization in the decision-making process
from the local rules. This process is shown in the
figure 4.

5. Applying Operators: Applying the operators
proposed in [24] for each rule group in this step.
6. Visualization: The visualization step is to
comprehend the results in an understandable form for
user actions.
7. Add to Contextitemsetlist\([N]\) two new candidates itemsets (ISC) 
   ISC\([N]\)(conAnt U SE) 
   ISC\([N]\)(conAnt U Cons U (RQ))
8. Else /*Local association rule*/
9. Add to Contextrulelist a new candidate rule 
   RC (con U Ant U SE \(\rightarrow\) Cons U (RQ-SE))
10. For each candidates itemsets 
    ISC\([N]\)\(\in\) Contextitemsetlist\([N]\) and a candidate rule 
    RC\(\in\) Contextrulelist
11. calculate the support s of itemsets in ISC\([N]\)
12. Verify the support s and confidence c for RC
13. Remove from a Contextrulelist all rules with 
    \(s<\text{minsup}\)
14. Return Contextrulelist and Contextitemsetlist

**Figure 5:** Auto-site Phase

5. Local Mining Algorithm: Apriori algorithm [25] is used 
   to discover patterns and relationships between attributes 
   from a given dataset. Apriori finds all frequent itemsets 
   that satisfies above a specified threshold values.
   CPAROR Auto-site phase extracts only interesting rules 
   that for the local user knowledge. In this Auto-site phase, 
   the search for interesting rules is performed locally and 
   associations between attributes that the user believes, 
   specified by means of the Context aware Rule Schemas. 
   It filters out those rules that are conform to user 
   knowledge and generating locally all candidate rules.
6. Forward the candidate rules to higher levels: these 
   candidate rules are important for finding the global rules 
   that are not frequent locally
7. Association Rules Visualization: The visualization step is 
   to comprehend the results in an understandable form for 
   the local user actions.

**CASE STUDY**
In this section, CPAROR framework is illustrated with a case 
studies, which considers college multiddatabase. We select 
students’ activity dataset for our case study. In the Figure 6 
describes the multi-level organization of the university. The 
upper level consists of the central unit, called University, and 
the lower level consist of departments and the middle level 
constitute the colleges under the university. All the 

- departments of the college have their own LMS database. The 
database consists of day-to-day students’ activity transaction 
log. Each department uses the same structural database 
given in the Figure 7.

- **Figure 6:** University Organization

- The experiment dataset was taken from the college LMS 
  multi-database consisting of 10, 367 items of students 
  transaction taken place in the odd semester for the year 2015-16.
  We chosen a specified set of threshold values for 
  dependent variable, support and confidence to mine 
  association rules based on our observation made on online 
  learning behavior of the students. The most popular activities 
  are viewing the course, completing feedback, adding and 
  updating posts and viewing discussions on a forum. As shown 
  in Figure 7, the number of student data set was used is 3, 140 
  records consisting of personal records, course (face-to-face) 
  records and students’ log file from the LMS.

- **Figure 7:** Student Log Details from LMS

Because of the unlimited discovered rules generated by the 
association rule mining process, we propose a new schema to 
represent user knowledge with context factors for the post-
mining process to reduce the number of rules a few.

**Context-aware Rule Schemas**
We proposed a set of context aware rule schemas for activity 
context like chat room, forum discussion, etc. with respect to 
slow learners and advanced learners. A context-aware rule 
schema allows to represent user expectations and context-
aware factors and permit the user to supervise association rule
mining process interactively. Meanwhile, auto-site and extern-
site operators applied over this schema to filter the discovered
rules. Table 1 specifies the context rule schema for the LMS.

Table 1: Context-aware Rule Schema

<table>
<thead>
<tr>
<th>Context Rule Schema Id</th>
<th>Context</th>
<th>Rule Schema</th>
<th>Operator</th>
</tr>
</thead>
<tbody>
<tr>
<td>CRS1</td>
<td>Spatial</td>
<td>&lt;student&gt;&lt;L&gt;&lt;1&gt;</td>
<td>Conforme</td>
</tr>
<tr>
<td>CRS2</td>
<td>User</td>
<td>&lt;course&gt;&lt;G&gt;&lt;3&gt;</td>
<td>Conforme</td>
</tr>
<tr>
<td>CRS3</td>
<td>Activity</td>
<td>&lt;student, course&gt;&lt;M&gt;&lt;2&gt;</td>
<td>Conforme</td>
</tr>
<tr>
<td>CRS4</td>
<td>Spatial</td>
<td>&lt;course-staff&gt;&lt;G&gt;&lt;2&gt;</td>
<td>Conforme</td>
</tr>
<tr>
<td>CRS5</td>
<td>Temporal</td>
<td>&lt;course-student&gt;&lt;E&gt;&lt;2&gt;</td>
<td>Unexpectedness Type</td>
</tr>
</tbody>
</table>

Context Aware Association Rules for Slow Learners

Table 2 shows the Context Association Rules for Slow Learners, i.e., from activity (X) to activity (Y). Students fill out the questionnaire, add discussions on a forum and then view discussions and search for the questions based on their interest. After adding a discussion on the forum, they view resources to find the answer to a subject question. The resources are the teaching material files, which are provided by the teacher. It is interesting that students in Slow Learners often remember to logout from the system and they pay more attention to the security of personal information than students in Advanced Learners. In the first case study, the analyst (decision maker at level branch 2) is interested in the identification of slow learners feedback activities in the LMS. The analyst already doesn’t know the direction of the implication and creates a Context-aware Rule Schema CRS1. The application of a conformed operator over this Rule Schema has the following majorities rules output: 62 results: Feedback → forum [S=24% C=60, 8%]

Feedback → Resource [S=21% C=59, 8%]

Table 2: Context Association Rule Schema for Slow Learners

<table>
<thead>
<tr>
<th>Activity (X)</th>
<th>Activity (Y)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Course</td>
<td>Recent</td>
</tr>
<tr>
<td>View</td>
<td>ImpliesUser</td>
</tr>
<tr>
<td>View</td>
<td>Update</td>
</tr>
<tr>
<td>Feedback</td>
<td>Complete</td>
</tr>
<tr>
<td>View</td>
<td>ImpliesForum</td>
</tr>
<tr>
<td></td>
<td>View forum</td>
</tr>
<tr>
<td>Feedback</td>
<td>Complete</td>
</tr>
<tr>
<td>View</td>
<td>ImpliesForum</td>
</tr>
<tr>
<td></td>
<td>View forum</td>
</tr>
<tr>
<td>Folder</td>
<td>View</td>
</tr>
<tr>
<td>Forum</td>
<td>Add discussion</td>
</tr>
<tr>
<td></td>
<td>ImpliesForum</td>
</tr>
<tr>
<td></td>
<td>View forum</td>
</tr>
<tr>
<td></td>
<td>ImpliesResource</td>
</tr>
<tr>
<td></td>
<td>View</td>
</tr>
</tbody>
</table>

Context Aware Association Rules for Advanced Learners

Table 3 shows the association rules between activities for Advanced Learners, i.e., from activity (X) to activity (Y). Students in Advanced Learners often view a uniform resource locator (URL) and resource after viewing the course and completing feedback. The URL is a web link to teaching videos on YouTube. Students in Advanced Learners review the lesson on the platform after it is taught. In addition, students in Advanced Learners keep in touch with teachers or classmates by writing messages. Students in Advanced Learners also often surf many websites at the same time and are more often compulsorily logged out of the after 5 minutes of no activity.

In the second case study, the analyst (decision maker at level branch 2) is interested on the identification of Advanced learners activities in the LMS. The analyst already doesn’t know the direction of the implication and creates a Context-aware Rule Schema CRS1. The application of a conformed operator over this Rule Schema has the following majorities rules output:
12 results: Feedback → forum [S=25% C=75, 8%]

Feedback → Resource [S=20% C=70, 8%]

Table 3: Context Association Rule Schema for Advanced Learners

<table>
<thead>
<tr>
<th>Activity (X)</th>
<th>Activity (Y)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Course</td>
<td>Recent</td>
</tr>
<tr>
<td>View</td>
<td>ImpliesUser</td>
</tr>
<tr>
<td>View</td>
<td>Login, Update</td>
</tr>
<tr>
<td>Feedback</td>
<td>Complete</td>
</tr>
<tr>
<td>View</td>
<td>ImpliesMessage</td>
</tr>
<tr>
<td></td>
<td>Write</td>
</tr>
<tr>
<td>Feedback</td>
<td>Complete</td>
</tr>
<tr>
<td>View</td>
<td>ImpliesMessage</td>
</tr>
<tr>
<td></td>
<td>Write</td>
</tr>
<tr>
<td>Folder</td>
<td>View</td>
</tr>
<tr>
<td>View</td>
<td>ImpliesUser</td>
</tr>
<tr>
<td></td>
<td>Login, update, view all</td>
</tr>
</tbody>
</table>
Z=(x-\bar{x})/\sigma_x

Where ‘x’ is the activity usage count, \bar{x} and \sigma_x are, respectively, the mean and standard deviation of the activity usage count and ‘Z’ is a semantic variable.

All of the continuous usage count values were normalized to discrete preference levels (PL) and 0.3 is selected to identify student learner type by using the semantic variable wit Z <-0. 3,-.0. 3 ≤ Z ≤ 0. 3, and z > 0. 3, to respectively represent inactive, neutral and active preferences. The preference level (PL) is the degree of preference that a student demonstrates for an activity, which is defined as follows, The preference level is 1 if the usage count, X > \bar{x} + 0.3 \sigma_x, and the preference score is -1, if the usage count, X > \bar{x} - 0.3\sigma_x, 0 represents a neutral preference. The student-activity preference score matrix is shown in the following Table 4.

<table>
<thead>
<tr>
<th>Activity Preference Scores for Slow Learners and Advanced Learners</th>
</tr>
</thead>
</table>

It is firstly necessary to identify the characteristics and levels of slow learners and advanced learners. The average usage count for activity X in a cluster is compared to the average usage count plus/minus the standard deviation, \sigma_x, for all students, to obtain the cluster preference score (PS), using the following formula

Z=(x-\bar{x})/\sigma_x

Table 2 and Table 3 show that students in both groups view other classmate’s recent profiles, which include information on the last login to the system. Students wish to know whether classmates view their recent posts and discussions on a forum, after they add posts and discussions on a forum. Students often check classmates’ latest login time to the system. They also often update their personal photos on the LMS. It is found that students care about the peer status in the collaborative environment. In addition, students often observe other students learning status and wish to attract their attention. When they add a new topic for discussion on a forum, they view the posts that other classmates add or update. At the same time, they regularly update their photos and personal information on their own home pages. They attract the attention of peers for personal and social reasons.

The preference score for an activity is PS(X_{type of Students, Activity}). For example, PS( X_{view course}^0)=-1 and PS( X_{view course}^1)=1. The activity preference scores of slow learners and advanced learners are shown in Table 5. The preference scores for activities such as viewing the course, completing feedback, adding posts, updating posts and viewing discussions on a forum in slow learners are PS( X_{view course}^0), PS( X_{complete feedback}^0), PS( X_{add post}^0), PS( X_{update post}^0), PS( X_{view discussion}^0) which is (-1,-1,-1,-1,1) and The preference scores for these activities for advanced learners are PS(X_{view course}^1), PS( X_{complete feedback}^1), PS( X_{add post}^1), PS( X_{update post}^1), PS( X_{view discussion}^1) which is (1, 1, 1, 1, 1)

Table 5: Preference Level for Activities of Slow and Advanced Learners

<table>
<thead>
<tr>
<th>Learner Type</th>
<th>Activity</th>
<th>X</th>
<th>\bar{x}</th>
<th>x - 0.3\sigma_x</th>
<th>x + 0.3\sigma_x</th>
<th>Preference Level (PL)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Slow Learners</td>
<td>Course</td>
<td>Recent</td>
<td>3.0</td>
<td>2.4</td>
<td>3.6</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>View</td>
<td>269.3</td>
<td>399.0</td>
<td>319.8</td>
<td>478.2</td>
<td>-1</td>
</tr>
<tr>
<td></td>
<td>Feedbac k</td>
<td>Complete</td>
<td>7.0</td>
<td>9.0</td>
<td>7.8</td>
<td>10.2</td>
</tr>
<tr>
<td></td>
<td>Folder</td>
<td>View</td>
<td>1.2</td>
<td>10.0</td>
<td>7.9</td>
<td>12.1</td>
</tr>
<tr>
<td></td>
<td>Forum</td>
<td>Add discussion</td>
<td>2.0</td>
<td>2.0</td>
<td>2.0</td>
<td>2.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Add post</td>
<td>38.3</td>
<td>53.0</td>
<td>44.3</td>
<td>61.7</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Delete post</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Update post</td>
<td>1.0</td>
<td>3.0</td>
<td>2.4</td>
<td>3.6</td>
</tr>
<tr>
<td></td>
<td></td>
<td>View discussion</td>
<td>294.3</td>
<td>402.0</td>
<td>332.7</td>
<td>471.3</td>
</tr>
<tr>
<td>Advanced Learners</td>
<td>Course</td>
<td>Recent</td>
<td>2.5</td>
<td>3.0</td>
<td>2.4</td>
<td>3.6</td>
</tr>
<tr>
<td></td>
<td>View</td>
<td>519.3</td>
<td>399.0</td>
<td>319.8</td>
<td>478.2</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Feedbac k</td>
<td>Complete</td>
<td>14.7</td>
<td>9.0</td>
<td>7.8</td>
<td>10.2</td>
</tr>
<tr>
<td></td>
<td>Folder</td>
<td>View</td>
<td>13.3</td>
<td>10.0</td>
<td>7.9</td>
<td>12.1</td>
</tr>
<tr>
<td></td>
<td>Forum</td>
<td>Add discussion</td>
<td>3.0</td>
<td>2.0</td>
<td>2.0</td>
<td>2.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Add post</td>
<td>78.5</td>
<td>53.0</td>
<td>44.3</td>
<td>61.7</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Delete post</td>
<td>1.5</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Update post</td>
<td>5.5</td>
<td>3.0</td>
<td>2.4</td>
<td>3.6</td>
</tr>
<tr>
<td></td>
<td></td>
<td>View discussion</td>
<td>531.3</td>
<td>402.0</td>
<td>332.7</td>
<td>471.3</td>
</tr>
</tbody>
</table>

Table 4: Student-Activity preference Score Matrix

<table>
<thead>
<tr>
<th>Student ID</th>
<th>View Courses</th>
<th>Complete Feedback</th>
<th>Add post on forum</th>
<th>Update post on forum</th>
<th>View discussion on forum</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>-1</td>
<td>0</td>
<td>-1</td>
<td>1</td>
</tr>
</tbody>
</table>
“Recent course” means that students check the latest course status on the LMS. “Viewing folders” means that students download the necessary files, e.g., teaching plans and material, which are placed in folders on the LMS. “Deleting posts” means that students delete their post after adding a post on a forum. The usage counts for these activities for Advanced Learners are twice those for Slow Learners. Slow Learners is an inactive group and Advanced Learners is an active group. Students in the inactive group (Slow Learners) seldom view the course, complete feedback, add posts, update posts, or view discussions on a forum. However, the active group (Advanced Learners) students have the opposite learning behaviors. Table 2 shows that the usage counts for adding posts, updating posts and adding discussions for Advanced Learners are all larger than those for Slow Learners. Students in Advanced Learners show confidence in self-expression. The usage counts for recent course, viewing folders and deleting posts for Slow Learners are all close to those for Advanced Learners. In summary, students in both groups download the necessary teaching material files, check the latest course information and delete their posts on a forum. The framework is tested through this case study with the live datasets in multi-database to make the model more efficient. The rules obtained by this study were useful for the administrators at different levels of college administrations to the effective decision making process.

CONCLUSION
This paper addresses the issue of selecting interesting association rules by analyzing the huge volumes of delivered rules in multi-data base mining process. For this purpose, we propose a new kind of formalism, which combines the user knowledge and context factors using ontology and Context-aware Rule Schema Multi-Level and a set of operators were applied over this schema to filter the rules. In auto-site phase, association rules are extracted by the mining process and forwarded to extern-site. In the extern-site phase, the forwarded rules from auto-site are synthesized and ranked them in order to select interesting rules and this process was driven by ontology and context-awareness. The case study described in this paper gives useful results for the decision-maker in educational domain. This paper is opened a door for further research in the following directions: Ubiquitous learning (U-learning), developing another new set of operators applied over the schema and applying this generic framework in medical and other domains also.

REFERENCES


