Sequential Rule Mining in M-Learning Domain

Shivesh Tiwari  
Assistant Professor  
Department of CS & IT Engineering  
B.B.S College of Engg. & Technology  
Allahabad, Uttar Pradesh, India-211013

Lokendra Kumar Tiwari, PhD  
Assistant Professor  
Indian Institute of Information Technology  
Deoghat, Jhalwa  
Allahabad, Uttar Pradesh, India-211012

ABSTRACT
Use of Sequential Rule mining is becoming an important tool in m-learning domain to convert the data into information. It is commonly used in a wide series of profiling practices, such as marketing, fraud detection and scientific discovery. Sequential Rule mining is the specialized technique using which we can extract some patterns from given data. These rules can be used to uncover patterns in data but is often performed only on given sample of data. The mining process will be ineffective if the samples are not good representation of the larger body of the data. The discovery of a particular pattern in a particular set of data does not necessarily mean that pattern is found elsewhere in the larger data from which that sample was drawn. An important part of the method is the verification and validation of patterns on other samples of data.

Since almost half of the world’s population are mobile phone holders; M-learning appears to be a promising field since it’s empowering learners with the ability to learn anytime and anywhere, access learning resources, communicate and interact with other learners all over the globe, participate in creating learning resources, and access the World Wide Web; thus many Open Universities are coming up with a new concept of the web which is mobile learning (m-learning); which is internet for small screen that would be suitable and viewable on a mobile devices which allow quick access to web content;

Sequential Rule mining is a emerging research area in the field of M-Learning; this concept of data mining in m-learning is still in its growing stage; thus this paper aims to present an approach for building a m-learning architecture, keeping in mind the server based content adaptation approach of backend database. Learners are made to create their profiles and their contextual information is stored at server end. This information is updated by learner, whenever learner makes an attempt to access learning content sequential role mining comes into major role to help extracting some behavioral patterns from the database.

General Terms
Clustering, pattern data, mobile learning, internet, database, behavioral patterns, world wide web

Keywords
Data mining, Sequential Rule Mining, m-Learning, sample data

1. INTRODUCTION
Nowadays the education industry have changed the concept and mode of learning from class room learning to e-learning and from e-learning to now what is known as m-learning or mobile learning. With the advent of the mobile device architecture revolution these handheld mobile devices with advanced features can perform several complex functions, but designing the learning content to fit the capabilities of mobile devices is a significant problem in different aspects. Mobile technology has made a remarkable contribution to the human life not only to communication and connectivity but it’s contribution has also been significant in education. Unfortunately, most of learning materials are designed for displaying on desktop computers that lead to unlikable presentation of most web pages on the small devices. Mobile devices needs delivered content to be adapted according to their device context and user context.

With the advent of new communication technologies, Traditional learning methods have now been transformed from “classroom” to “virtual-learning” to “smart-learning”; Virtual learning is a part of E-learning methodology where a user obtains learning experience through automated teaching and learning processes thereby deploying Expert Systems, or Knowledge Databases in combination with web-based technologies for universities and other educational organizations.

The term m-learning stands for mobile learning, a relatively new medium of education, whereby the educational content delivery and uptake occur “on the move”, typically outside the standard learning environment of classroom and home. A m-learning can be thought of as an extension to e-learning with added feature of mobility component. Its functioning is based on the paradigm of “anytime, anywhere”. A m-learning is enabled by integrating various hardware and software technologies into multimedia application, facilitating the communication of educational content in number of different formats. E.g. quizzes, games. The end user element of most m-learning application involves a mobile device (which can be of different type, e.g. cellular phone, iPod, smart phone etc.). The applications are usually supported by latest communication infrastructure e.g. 3G, 4G, GPRS and Wi-Fi.

The successful proliferation of m-learning requires a sustained activity to understand needs, wants and limitations of potential user of m-learning application. A student of any institution may belong to different backgrounds. They have different living environment, family background and their mobile device type may differ from each other in many ways.

The factors which could affect a m-learning application working can be broadly classified in two parts:

1. Device Context: A learner may be using various types of devices. The mobile application should be designed in such a way that heterogeneity of these mobile devices doesn’t hamper its functioning. The different features of mobile devices that would be different from one another and may pose restriction on m-learning working are: Device screen size; Screen resolution; Processing, storage and battery capacity of device; Type of data format, a device would execute; Type of connection a device is using etc.
2. Learner’s Context: A m-learning application’s user should be foremost concern of any developed application. The preferences of learner should be considered at highest priority. The learner’s living environment can differ from user to user and changes dynamically. Their life schedule, surrounding environment plays vital role in deciding the preferences of learner. Few important learner’s context which must be included in any m-learning application are: Type of data format in which content needs to be received; Noise level and luminosity level of learner’s surrounding; Learner is moving or stationary etc. These contexts may change dynamically from time to time and learning application needs to adapt these changes accordingly.

This research paper attempts to present an approach for building a m-learning architecture, keeping in mind above mentioned requirements. Here the architecture follows server based content adaptation approach. Learners are made to create their profiles and their contextual information is stored at server end. This information is updated by learner, whenever learner makes an attempt to access learning content.

2. M-LEARNING
One definition of mobile learning is “Any sort of learning that happens when the learner is not at a fixed, predetermined location, or learning that happens when the learner takes advantage of the learning opportunities offered by mobile technologies.”

The main characteristics of m-learning are:

1) Ubiquitous and on Demand: Ubiquitous means, which is present everywhere. Hence it should be accessible regardless of time and location and capable of delivering the required content at any point of need.

2) Bit Sized: The educational content should be relatively short in duration as it is used mostly at places having environment causing a lot of distraction.

3) Blended: It is very rarely as only platform to deliver education. Normally it complements other more resourceful modes of content delivery such as clinical and e-learning.

4) Can be Collaborative: It should take advantage of mobile communication devices it uses as its basis (e.g. mobile phones, Wi-Fi enabled PDA.) and promote collaborative learning as much as possible

3. FACTORS INVOLVED IN M-LEARNING
While designing mobile learning environment, following factors must be considered:

1) Device Context: While delivering content, the configuration of end mobile device must be known. There is variety of mobile devices present in market, each having different screen size, environment, processing and storage capability, battery capacity etc. The delivering content must be adapted according to user’s device.

2) User’s Context: The user of mobile learning must have the privilege to choose the way in which content should be presented. The content to be delivered can be in text, audio or video format. Factors like location of user (i.e. user is in static or dynamic mode); noise level and brightness level of the user’s place should be considered.

3) Transmission Time: The time which mobile content takes to be delivered from learning server to the end mobile device is important design issue. Higher transmission time will result in increased data costs for user. Hence this factor holds great importance.

4) Different Devices using Different Platforms: While designing any mobile application, different platform supported by different mobiles needs to be considered. For example Android phone supports Java, C++; iPhone, Windows phone supports C++ and Symbian phones supports Java. The application to be designed should run in all kinds of platforms of various kinds of mobile devices.

4. SEQUENTIAL RULE MINING
In today’s world of fast growing economy, businesses and daily transactions, the need to store huge amount of data and thereby transforming those bulky data into usable information is becoming key research area in various domains including databases, machine learning, statistics, and so on. As far as databases are concerned, the researchers now days are mainly focused about dealing with finding the patterns of occurrences of data, texts, images, transaction records. All these efforts are only done to draw useful patterns hidden behind the data stored in back end. In this research paper our aim, is to develop and design some efficient and scalable rules for extracting a pattern to design an algorithms as integrated tools for m-learning domain.

Many researchers have defined the term Data mining, as per their convenience and also classified its major components as per their requirements. In general definition we can relate knowledge discovery and data mining in databases, interrelated to each other. Knowledge discovery can be recognized as the process of extracting non-trivial, implicit, previously unknown, and potentially useful information from data stored in small data stores, or huge databases in businesses or the data stored in portable devices like mobile phones or smart phones. The database used in the mining process generally contains large amounts of data collected by computerized applications through retail stores, digital sensors in scientific experiments, or through any automation tool used in research work. These databases thus serve as rich and reliable sources for knowledge generation and verification. The extracted knowledge from these databases can be used in many ways in corresponding applications like m-learning domain. Discovering patterns of customer browsing and transactions may assist the modeling of user behaviors for customer retention or personalized services. Given the desired databases, whether relational, transactional, spatial, temporal, or multimedia ones, we may obtain useful information after
A typical process of knowledge discovery in databases is illustrated in Fig. 1.

Fig. 1. The process of knowledge discovery in databases

the knowledge discovery process if appropriate mining techniques are used.

With the available Databases, initially we need to have significant and all round aspects of the knowledge about the stored data like about the data warehouse, transaction database and the goals of the application domain, along with the target data thereby setting and identifying the goals. New target data is created by selecting the data required from any of the linked data warehouses or transaction database or flat file source whatever applicable. The data cleaning process in Fig. 1 shown above is done as part of data preprocessing is used to remove the ‘dirty’ data, e.g. data with incomplete fields, missing or wrong values, in the preprocessing stage. The ‘clean’ data is then reduced and/or transformed so that the data is represented by the sequential rules applied for useful features and as part of data transformations. To find the useful patterns, the user is required to apply the mining rules, which include summarization/generalization of data characteristics, classification/clustering of data for future prediction, association finding for data correlation, trend and evolution analysis. The discovered patterns are evaluated and presented as knowledge. The process may iterate and contain certain loops between any two steps.

Finding all the frequent patterns from the huge data sets is a very time-consuming task. Although the frequency of a pattern can be determined by scanning the database once, the elements of the pattern cannot be known in advance. Take association discovery for example. Given 100 distinct items in the database, the total number of potentially frequent sets is \( C(100, 1) + C(100, 2) + C(100, 3) + \ldots + C(100, 99) + C(100, 100) \), where \( C(m, n) \) represents the combinations to choose \( n \) items from \( m \) distinct items. The total number of potential patterns is too huge so that validating all the potential patterns in a single database scanning could be impossible. Thus, it is desirable to design efficient algorithms for frequent pattern mining.

In addition, the mining algorithm must be scalable to handle databases of huge size. While the response time may be tolerable for an algorithm to check thousands of potential patterns against a small database having thousands of records, it could be intolerable against a database having millions of records. Similarly, an algorithm that assumes the database has maximum 100 elements might fail to mine any database having more than 100 elements. In the mining of frequent patterns in database context, the number of elements and the size of the database could be very large. Any improper assumptions on database or main memory could possibly produce an impractical algorithm that works well for small problems only.

We will propose a novel algorithm for mining sequential rules common to several sequences. Unlike other algorithms, new algorithm uses a pattern-growth approach for discovering sequential rules such that it can be much more efficient and scalable. The proposed algorithm will outperform CMRules and CMDeo in terms of execution time and memory usage.

4.1 Algorithm

The algorithm that we propose uses an approach that is different from CMDeo and CMRules. Instead of using a generate-candidate-and-test approach, it relies on a Pattern-Growth approach similar to the one used in the PrefixSpan \[7\] algorithm for sequential pattern mining. Our algorithm first find rules between two items and then recursively grow them by scanning the database for single items that could expand their left or right parts (these processes are called left and right expansions). Like PrefixSpan, Our algorithm also includes some ideas to prevent scanning the whole database every time. The idea of proposed algorithm is to grow rule by starting with rules of size 1*1 and to recursively add one item at a time to the left or right side of a rule (left/right expansions) to find larger rules.
PROPOSED MODEL

5. PROBLEM STATEMENT

The problem addressed in this research paper has two broader parts. Firstly, study and analysis of mobile learning and its issues which arises during its design phase is done. Secondly, architecture is proposed which enables delivery of learning content according to mobile device’s context and user’s context using sequential rule mining approaches.

6. PROPOSED MODEL

A m-learning program should focus on these things primarily:

- Content should be available anytime, anywhere. Internet connections may not be available at certain point of time, to deal with that, strategies should be taken so that offline contents are available. OCC applications (Occasionally Connected Computing, a term coined by ADOBE) can cache data to client and function even when internet connection is not present. An OCC solution is Google Gears- a browse extension that provides an API to run applications offline. It has a caching server, an offline SQL database.

- Content adaptation strategy should meet the goals of presenting materials to the client in such a way where cognitive gain is optimum. To do that transforming actual content dynamically as and when requested is necessary.

- WALI or Wireless Abstraction Library This is an open source tag library that transforms the WALE tags into WML, XHTML code.

- W3C’s DIAL (Device Independent Authoring Language) is a language that allows content delivery across devices and contexts. Using Wall or DIAL it is possible to dynamically resize images or generate markup codes dynamically.

- MyMobileWeb is java based open source framework. It is working on semantic mobile web.

In order to be successfully implementing, m-learning infrastructure, m-learning service needs to be:

- Highly compatible with „Mobile ,lifestyle (by offering specific mobile value).
- Student-centered and providing a clear learning value.
- Of high technical quality.
Mobile context can be defined as contextual information such as people, technological artifacts, learner location and its physical conditions, available time etc. Learning context can be described by elements of a learning design. A learning design is description of teaching learning process, which follows, a specific pedagogical model that places in unit of learning.

A sequential rule typically has the form X⇒Y. A sequential rule X⇒Y has two properties:

- **Support**: the number of sequences where X occurs before Y, divided by the number of sequences.
- **Confidence**: the number of sequences where X occurs before Y, divided by the number of sequences where X occurs.

Sequential Rule Mining finds all valid rules, rules with a support and confidence not less than user-defined thresholds minSup and minConf. **For Example**: An example of Sequential Rule Mining is as follows: Consider minSup= 0.5 and minConf= 0.5:

<table>
<thead>
<tr>
<th>ID</th>
<th>Sequences</th>
</tr>
</thead>
<tbody>
<tr>
<td>Seq1</td>
<td>{a,b},{c},{d},{f},{g},{e}</td>
</tr>
<tr>
<td>Seq2</td>
<td>{a,d},{c},{b},{a,b,e,f}</td>
</tr>
<tr>
<td>Seq3</td>
<td>{a},{b},{f},{c},{e}</td>
</tr>
<tr>
<td>Seq4</td>
<td>{b},{f,g}</td>
</tr>
</tbody>
</table>

**Fig 2: A sequence database**

<table>
<thead>
<tr>
<th>ID</th>
<th>Rule</th>
<th>Support</th>
<th>Confidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>r1</td>
<td>{a,b,c} ⇒ {e}</td>
<td>0.5</td>
<td>1.0</td>
</tr>
<tr>
<td>r2</td>
<td>{a} ⇒ {c,e,f}</td>
<td>0.5</td>
<td>0.66</td>
</tr>
<tr>
<td>r3</td>
<td>{a,b} ⇒ {e,f}</td>
<td>0.5</td>
<td>1.0</td>
</tr>
<tr>
<td>r4</td>
<td>{b} ⇒ {e,f}</td>
<td>0.75</td>
<td>0.75</td>
</tr>
<tr>
<td>r5</td>
<td>{a} ⇒ {e,f}</td>
<td>0.75</td>
<td>1.0</td>
</tr>
<tr>
<td>r6</td>
<td>{c} ⇒ {f}</td>
<td>0.5</td>
<td>1.0</td>
</tr>
<tr>
<td>r7</td>
<td>{a} ⇒ {b}</td>
<td>0.5</td>
<td>0.66</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>

**Fig 3: some rules found**

It uses a pattern-growth approach for discovering valid rules such that it can be much more efficient and scalable. It first finds rules between two items and then recursively grows them by scanning the database for single items that could expand their left or right parts. We have evaluated the performance of our algorithm by comparing it with the CMDeo and CMRules algorithms. Results show that our algorithm clearly outperforms CMRules and CMDeo and has a better scalability.

This model includes following steps which can be explained as follows.

1) Select the m-learning database or target subset of the database on which the data mining tools will be applied; further it is necessary to identify the goal of the mining from different perspectives.
2) Device and Learner’s Context

3) Cleaning and processing data which include handling errors and noisy data, remove redundancy, managing missing data, and preparing database for the mining process.

4) Clustering m-learners into similar groups based on factors that must be defined earlier depending on the learning provider point of view, these factors might be m-learners backgrounds, gender, major, common interests, etc.

5) Applying various data mining tools such as association rules, classification, genetic algorithms and other techniques to predict m-learners behavioral and thus identify m-learners patterns.

**6.1 Working of m-learning tool**

The m-learning tool is basically made up of four components, shown in Figure 4:

1. **Mobile Client Browser**: When a learner wants to access learning content, stored at the web server, the learner needs to log on to the corresponding account. The learner uses default phone browser and types server address on it. This in turn opens login screen on the phone. On submitting login ID and password, phone browser invokes learning web server to respond.

2. **Web Server**: A request from mobile client browser arrives at the web server. Upon receiving the request, web server invokes the corresponding web services. Then it acts according to web service response e.g. when web service tells the preferences of modes of data access and the corresponding mobile device i.e. User preferences of modes of data access, type of mobile

3. **Web Services**: Upon receiving any request, web server invokes corresponding web services. For each task separate web service is designed and in each action the web service accesses the database tables stored in the database. It sends query to database for creating, updating and checking records as per the sequential rules and algorithms designed in the database.

4. **Profile Database**: Each m-learner profile is created during the registration process and stored at UserInfo Table. This database stores records of learners who have joined the course. These records contain contextual information related to the student and the corresponding mobile device i.e. User preferences of modes of data access, type of mobile
phone, connection type, noise level, luminosity level etc.

which suits user. In this process, transmission cost is reduced significantly, as files to be sent are pre converted to their different formats.

8. REFERENCES


Management, Kansas, Missouri, USA, pp. 251-258, Nov. 1999.


[32]. RamuParupalli,Sarat Chandra BabuNelaturu, Dhanader Kumar Jain, “The Role of Content Adaptation in Ubiquitous Learning”, IEEE International Conference on Technology for Education.

