

## An Educational Data Mining Approach to Concept Map Construction for Web based Learning

Anal ACHARYA<sup>1</sup>, Devadatta SINHA<sup>2</sup>

<sup>1</sup>St. Xavier's College, Kolkata, India

<sup>2</sup>University of Calcutta, Kolkata, India

anal\_acharya@yahoo.com, devadatta.sinha@gmail.com

*This aim of this article is to study the use of Educational Data Mining (EDM) techniques in constructing concept maps for organizing knowledge in web based learning systems whereby studying their synergistic effects in enhancing learning. This article first provides a tutorial based introduction to EDM. The applicability of web based learning systems in enhancing the efficiency of EDM techniques in real time environment is investigated. Web based learning systems often use a tool for organizing knowledge. This article explores the use of one such tool called concept map for this purpose. The pioneering works by various researchers who proposed web based learning systems in personalized and collaborative environment in this arena are next presented. A set of parameters are proposed based on which personalized and collaborative learning applications may be generalized and their performances compared. It is found that personalized learning environment uses EDM techniques more exhaustively compared to collaborative learning for concept map construction in web based environment. This article can be used as a starting point for freshers who would like to use EDM techniques for concept map construction for web based learning purposes.*

**Keywords:** Educational Data Mining (EDM), Web based Learning, Concept Map, Personalized Learning, Collaborative Learning.

### 1 Introduction

The volume of data stored in the database is increasing at a tremendous speed. For this purpose tools and techniques are required to help humans automatically analyze this data set for useful information. This gave birth to a new field called data mining. Data mining has attracted attention from researchers from different fields including database systems, statistics, machine learning and pattern matching. The aim of this article is to study the use of data mining techniques for extracting interesting and useful patterns from web based learning systems and hereby improve the quality of learning.

Data Mining is the process of uncovering interesting data patterns hidden in large data sets [1]. It has been in use for many years in activities such as commerce and business along with host of government, public and private sector enterprises to gain meaningful insight into huge data sets generated by them. There are several other terms which have meaning similar to data mining such as knowledge mining from data, knowledge

extraction, and data/knowledge analysis. The term Knowledge Discovery from Databases (KDD) is often used in this context. Most researchers view data mining as an essential step in the process of KDD. The iterative nature of the KDD process is shown in Fig. 1. It consists of the following steps [2]:

- *Data cleaning*: used to remove noise.
- *Data integration*: used to combine data from multiple sources.
- *Data selection*: used to retrieve data relevant to analysis from the database.
- *Data transformation*: used for transforming data into forms appropriate for mining by performing summary operations.
- *Data mining*: used to apply methods for retrieving data patterns.
- *Pattern evaluation*: uses certain measures to evaluate interestingness of derived patterns.
- *Knowledge presentation*: used to present mined knowledge to user.

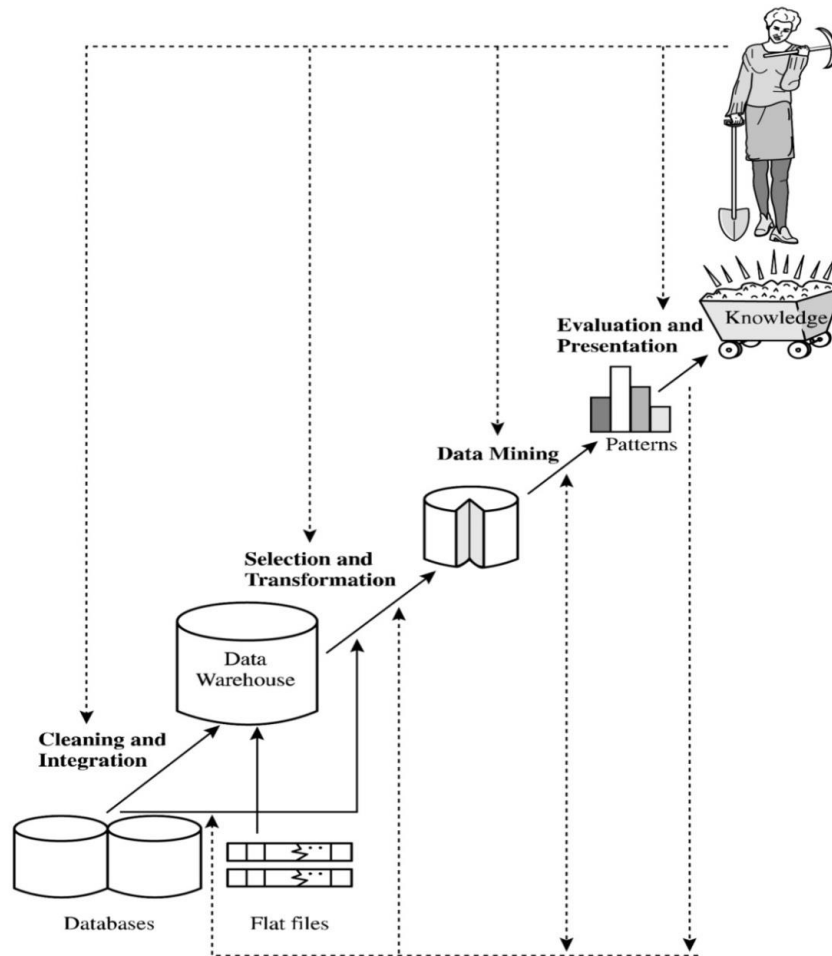


Fig. 1. Steps of the KDD Process [2]

Various data mining algorithms are used in context of learning systems for this purpose:

- *Classification*: The aim of classification is to separate items into several predefined classes. A collection of training samples is first assembled. A classification model is then designed to find a model for class attributes as a function of the values of other attributes [3].
- *Clustering*: The process of grouping a set of physical or abstract objects into classes of similar objects is called clustering. A cluster is a collection of data objects that are similar to one another within the same cluster and are dissimilar to the objects in other clusters [4].
- *Association rule analysis*: Frequent patterns are patterns (such as itemsets, subsequences, or substructures) that

appear in a data set frequently. Finding such frequent patterns plays an essential role in mining associations, correlations, and many other interesting relationships among data [2].

- *Regression analysis*: Data can be smoothed by fitting the data to a function, such as with regression. It can also be used for predicting a value of a given continuously valued variable based on the values of other variables. Linear regression involves finding the “best” line to fit two attributes (or variables), so that one attribute can be used to predict the other [1].

In this background the article is divided into following sections. The next section introduces the use of data mining in education along with various entities involved in the process and their functionalities. The

collection of techniques used for this purpose is called Educational Data Mining (EDM). A brief survey of EDM tasks already performed in personalized and collaborative environments is also presented. Web based learning systems may be used for enhancing the efficacy of EDM techniques. Section 3 provides a brief justification on that count. Web based learning systems often uses a tool for organizing knowledge. One such tool studied in this article is concept map. Section 4 discusses the applicability of concept maps in organizing knowledge in web based learning systems. Section 5 presents the web learning systems developed on the basis of EDM techniques which use concept maps for knowledge organization. Section 6 attempts to infer the general features of EDM techniques used for constructing concept maps in personalized and collaborative environments and compare their performances. Section 7 presents the limitations of the current works. The article concludes by providing certain pathways that may be explored by the future researchers.

## 2 What is Educational Data Mining?

Educational Data Mining (EDM) is the application of data mining techniques to educational data for its analysis [5]. It is thus concerned with implementing data mining algorithms for finding unknown student learning patterns. In other words, they help the learner and the instructor in improving the teaching learning process. There are two perspectives to educational data mining process: Firstly increased automation of educational institutions creates large repositories of student data. Secondly, as mentioned above, the use of internet in education means that this dataset can be accessed in a real time fashion. EDM is thus concerned with developing methods to explore large datasets in educational environment for better understanding of student learning patterns resulting in the benefit of the learners [6].

There are several entities involved in the EDM process. The advantages gained by them are next discussed:

(i) *Learners*: The data mining process recommends activities and resources to the learner so that they may further improve their learning. It also identifies interesting learning activities which could be used to identify optimal learning paths. Further, courses, discussions and books could be identified as a result of this activity.

(ii) *Instructors*: The teachers can get a feedback about the courses offered by them instantaneously. They can also detect student learning patterns and their nature of mistakes. Learner historical record could be used to predict their performance in examinations. This in turn could be used for identifying where they need additional learning support.

(iii) *Course developers*: Their main aim is to create, maintain and evaluate courseware. Data mining process could be used to improve the quality of courseware based on student learning experiences. They could also be used to develop student and learner models in an automated fashion.

(iv) *Academic administrators*: Their aim is identify the courses that would be suitable to a class of learners. Data mining process could be used in selecting the best candidates for admission by identifying who would do well in the later stages of learning. Further they can help in finding the most effective way of improving student grades.

(v) *System administrators*: They are involved in looking after the learning systems. Data mining process could be used to organize the institution's resources in an optimal fashion so that they can be utilized more effectively. They can also be used to enhance the quality of the learning systems.

The entities that stand to gain out of the data mining process were previously discussed. The important educational tasks performed by these entities using data mining process and their use in various types of learning environments are next discussed:

(i) *Analysis of student learning patterns*: The objective of this analysis is to study student learning activities to get useful information for decision making. Statistical algorithms are mostly used for this purpose. When applied on educational data, they can provide summary

reports about learner behavior. Software such as Statistical Package for Social Sciences (SPSS) has commonly been used for this purpose.

(ii) *Providing mechanisms for betterment of learning activity*: The main aim here is to identify interesting and uncommon learning traits among the students so that instructors and academic administrators may use these for betterment of learning. Classification, clustering and association rule mining have been commonly used for this purpose.

(iii) *Identifying future performance of a learner*: The objective here is to predict the performance of a student based on his previous learning records. The value normally predicted is performance or score. Generally two types of algorithms are used for this purpose: regression analysis and classification.

(iv) *Grouping learners based on certain traits*: The objective here is to group learners based upon certain characteristics like personal working style, subject knowledge etc. The group of learners generated can be used for developing personalized learning system or this group of students can learn collaboratively. Classification and clustering algorithms are used for group formation purpose.

(v) *Development of learning objects*: The purpose of constructing courseware is to help learners learn the subject in an easy and automated fashion. It should also promote learning object reuse among several learners. Clustering and hybrid data mining techniques have been used for this purpose.

(vi) *Using tools for organizing knowledge*: A class of tools called mind tools is used to organize knowledge in a hierarchical fashion. Association rule mining has generally been used to construct these tools in an automated fashion. This article explores the use of one such tool called concept map for organizing knowledge.

Learning systems can be classified into two categories based on the nature of applications: Personalized Learning (P-Learning) and Collaborative Learning (C-Learning). In P-learning, each student can plan his curriculum to meet his needs [7, 8, 9]. Again several learners may attempt to learn something together capitalizing on one another's resources and skills resulting in increased learning efficacy leading to C-Learning [10, 11]. Some examples of EDM tasks discussed above in P-Learning and C-Learning environment are enumerated in Table 1.

**Table 1.** Examples of P-Learning and C-Learning applications using EDM in referred task domains

Task Domain	P-Learning	C-Learning
Analysis of student learning patterns	Most popular web pages used by the learner [12]. Number of pages browsed by the learner [13]. Patterns of users through various time periods [14].	Statistical patterns of learner interaction in a group [15]. Overall contribution of a learner in a group [16]. Amount of posting vs. replies in a group [17].
Providing mechanisms for betterment of learning activity	Whether resources used by a student have a positive impact on performance [18]. To discover association patterns between various students attributes [19].	Determine relationship between each learning behavior pattern [20]. Enable teachers to analyze, refine and reorganize reading materials [21].
Identifying future performance of a learner	Identifying students who may perform poorly in term end exams [22].	Predict future performance of a group [24].

	Predicting time spent on learning a page [23].	Recommendation model for students in similar situations in the future [25].
Grouping learners based on certain traits	Group students based on the contents of web page they are visiting [26]. Interesting patterns characterizing the strong and weak students [27].	Group students according to their individual learning styles [28]. Identify skills that differentiate students in a group [29].
Development of learning objects	Web tutor tree [30]. Learning objects reuse and retrieval from the web [31].	Profile analysis for searching and ranking learning objects based on predicted user interest [32].
Using tools for organizing knowledge	Construction of learning paths based on students historical test records [33] Identifying learning barrier and overcoming it [34].	Learning by constructing a learning sequence collaboratively and comparing it with the instructors [35].

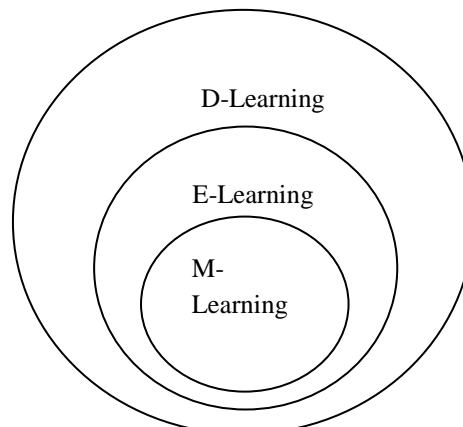
### 3 Applicability of EDM in Web based learning

Based on the above discussion two aspects of thought emerge. Firstly, learning systems generates a lot of student data which are stored in the servers of these systems. Analysis of these data can lead to uncovering a lot of interesting student learning patterns. The size of this data set is dependent on the number of learners. More the number of learners involved in the learning process, larger is this data set. For this purpose data mining algorithms need to be applied on this data set. Since the volume of data is quite large, some real time application is necessary so that results of the data mining process can be used to take corrective action. Secondly, most of the educators who are supposed to use the results of EDM process are not conversant with the process of data mining. They would like the entire process to be presented in the form of a tool to them so that they can access it in the form of an application and get the desired result. This would lead to EDM tools more widely used by the educators so that the feedback obtained from these could again be used in the EDM process. *Several researchers have proposed that a web based system would be a solution to both these problems.* In a web based system, the data stored in the web logs obtained due to interaction between the

learners, instructors and administrators could be accessed instantaneously as soon as it is generated. On applying data mining algorithms to these, the results obtained could immediately be given to the stakeholders for taking corrective actions. Also web based learning systems could be devised in such a manner so that the data mining tasks (preprocessing, data mining, post processing) could be accessed in the form of a single application through a simple interface. The roots of web based learning can be traced to Distance Learning (D-Learning). D-Learning is a general term used to cover the broad range of teaching and learning events in which the student is separated (at a distance) from the instructor, or other fellow learners [36]. As D-Learning began to rely more on technology, web based learning systems evolved. Web based learning is implemented in two ways. Electronic Learning (E-Learning) is used for learning from any device dependent upon the actions of electronics, such as television, computers, videodiscs, video games, radio, videotexts, teletext, and all the other devices in the process of being invented that are electronic in nature [37]. A major disadvantage of E-learning is its lack of ubiquity and idle time utilization. Learners would often like to utilize their time when they are on the move. The advancement of

mobile communications in late 90's gave birth to devices like Personal Digital Assistants (PDA), and smart-phones leading to a new type of learning called Mobile Learning (M-Learning). M-Learning is the intersection of mobile computing (the application of small, portable, and wireless computing and communication devices) and E-learning (learning facilitated and supported through the use of information and communications technology) [38]. M-Learning enables students to take courses at their convenience: (i) at a location they desire, (ii) at a time they choose, (iii) undisturbed, and (iv) facilitated by inexpensive, ubiquitous mobile devices. The relationship between these three forms of learning is shown in Figure 2.

Web based learning systems relies heavily on using a mind tools for knowledge organization. However for learning process to be efficient such tools should be generated in an automated fashion. As indicated earlier, web based learning systems generate a large volume of learner data which are stored in their web servers [3]. EDM techniques have often been used on this huge data set for finding the patterns of system usage by learners and discovering their learning behavior patterns. The techniques that have been employed for this purpose include clustering, classification, and association rule analysis along with a host of statistical, machine learning algorithms. This article explores the effect of one such tool called concept map in web based learning systems. A concept map [39] is a conceptual graph that shows relationships between concepts and expresses the hierarchal structure of knowledge. Concept maps are used for diagnosing the learning problems of students. In personalized environment automatically constructed concept maps will enable instructors to overcome the learning barrier and misconceptions among the learners whereas in collaborative environment learners discuss and construct concept maps for learning a subject. The next section introduces concept maps and its motivation in using as a tool in knowledge organization.



**Fig. 2.** Evolution of Web based Learning [39]

#### 4 Concept Maps as knowledge organization tool in Web Based Learning Systems

In recent times, various types of computer-based learning systems have been proposed by researchers for different types of learners in personalized as well as collaborative environments. Among them one that has attracted a lot of attention is Intelligent Tutoring System (ITS). An ITS is a computer system that aims to provide immediate and customized instruction and feedback to the learners without intervention by a human tutor [40]. The first major contribution in this area has been the work of Johnson [41] in which he constructed a authoring environment for building ITS for technical training for IBM-AT class of computers. Another contribution has been the work of Vasandani [42] where he built an ITS to organize system knowledge and operational information to enhance operator performance. There has been a lot of work in ITS in M-learning environments as well. In 2005, Virvou et al. [43] implemented a mobile authoring tool which was called Mobile Author. Once the tutoring system is created it can be used by the learners to access learning objects and tests. Around the same time, Kazi [44] proposed Voca Test, which is an ITS that helps to learn vocabulary using M-Learning approach.

However, ITS' suffers from two major disadvantages: First, they do not any use mind tools [35] to organize knowledge. A mind tool is a computer-based knowledge construction tool that enables the learner to organize the subject they are studying. Secondly and more

important, they do not provide immediate and specific feedback about a learner’s problems to a tutor [45]. This is due to the lack of face-to-face interaction between the learner and the tutor. This problem can be overcome by analyzing student learning behavior and closely monitoring learning deficiencies. Thus, the learning system should be able to generate some form of learning aid for the learners in a timely fashion.

Concept maps have often been used as a tool for organizing knowledge in web based learning systems. Concept maps were developed from the theory of meaningful learning proposed by David Ausubel [46] in 1963. In meaningful learning theory, the learner is able to relate the new knowledge to a relevant concept that he already knows. On the other hand, if a person attempts to learn by memorizing some information, he attempts rote learning. Ausubel thus advocated the concept of reception learning, where new concepts are obtained by adding questions and clarifications to old concepts. For this, he proposed two methods: signaling, which indicates important information, and advanced organizers, which indicate the sequence between them. These psychological foundations led to the development of concept maps by Joseph D. Novak at Cornell University in 1972. Since then, concept maps have been used by a lot of researchers to diagnose learning problems of students. In brief, let C1, C2, C3, C4 and C5 be five concepts as shown in Figure 3.

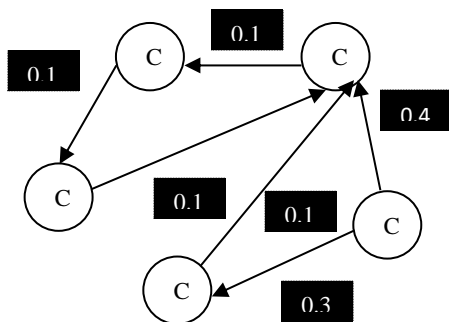


Fig. 3. A Concept Map of learning

The edge  $C1 \rightarrow C3$  is an association rule and indicates that the concept C1 should be learnt before concept C3 [34]. Similarly, learning concept C3 is a prerequisite to learning C2. In

other words, if a student fails to learn C3, it is due to the failure of mastery over concepts C1. There exists a confidence level 0.45 associated with the rule  $C1 \rightarrow C3$ , which states that if the student fails to understand C1, then the probability for him failing to understand C3 is 0.45 [2].

**5 Concept map based Web Learning Systems Developed using EDM Techniques**

The previous section leads to a position where a detailed discussion of the web learning systems developed using concept maps in web based environment using EDM techniques can be presented. Discussion has been done on two axes. Articles are first classified based on the nature of applications (personalized and collaborative). Within themselves they are further classified based on environment (E-Learning and M-Learning). Jong et. al. [47] suggests two main applications of concept maps. In the first type of applications, students learn a subject and construct concept maps based on it. The instructor then constructs his own concept map and compares it with the student’s to find his learning deficiencies. He named this method concept mapping. In second application, named diagnostic concept graph, a concept map is constructed and the student is asked to evaluate his learning status from it.

**5.1. Concept Maps in Personalized Applications**

It has been found that personalized applications are mostly developed using diagnostic concept maps. Current research [48] suggests that there are three methods of generating diagnostic concept maps: manual, semi automated and automated. To make the learning process effective and appealing to the students, automated construction of concept maps is necessary. In E-Learning environment, diagnostic concept maps were mostly constructed from students historical test records. Data mining heuristics were applied to these test records for generating concept maps. This process however cannot be followed in M-Learning systems due to the lack of sufficient computing power of mobile devices. This section first discusses the use of

concept maps in E-Learning environment and then in M-Learning environment.

Concept maps were first used by Hwang [33, 48] to generate learning guidance for students in science courses. In his work he created a Concept Effect Relationship (CER) builder which was used for automated construction of concept maps. He tested the system with physics, mathematics and natural science courses and proved statistically that CER indeed provides better learning diagnosis than traditional ITS. Lee et al. [34] in their work has used the Apriori algorithm to generate concept maps which has been used to generate learning guidance. The system built by them is called Intelligent Concept Diagnostic System (ICDS). ICDS generates the Remedial Instruction Path (RIP) for providing proper learning guidance. Tseng et. al. [49] has proposed a Two-Phase Concept Map Construction (TP-CMC) algorithm. Phase 1 is used to mine association rules from input data whereas phase 2 uses these association rules for creating concept maps. They also developed a prototype system of TP-CMC and used real testing records of students in a junior school to evaluate the results. The experimental results showed that TP-CMC approach works satisfactorily. An agent based system was developed by Chen et al. [50] to generate learning guidance. This system again used the Apriori algorithm to generate association rules. The system developed was named Personalized E-Learning System (PELS). It was used for diagnosing common misconceptions for a course on Fractions. Jong et al. [47] has devised a learning system using Sequential Probability Ratio Test (SPRT) to understand whether a student has learned a concept or not. Their study devised the 'Remedial-Instruction Decisive path (RID path)' algorithm for diagnosing individual student learning situation. Weights were assigned to the relations between the concepts to find out the missing concepts. They applied this method to two courses namely 'Introduction and Implementation of RS-232' and 'Electronic Circuits Laboratory' and found that students were satisfied with the remediation mechanism. Kohonen's self-

organizing map algorithm has been used by Hagiwara [51] to generate Self Organizing Concept Maps (SOCOMs). Computer simulation results done by him have shown the effectiveness of the proposed SOCOM. Ling-Hsiu Chen [52] in his work has used pathfinder networks which is an advanced form of concept maps for providing remedial learning paths for individual learners based on their knowledge structure. He named the system Personalized Diagnosis and Remedial Learning System (PDRLS). He tested the system with 145 students enrolled in introductory Java programming language courses at a Central Taiwan technology university. His experimental results demonstrated that learners who received personalized remedial learning guidance via PDRLS achieved improved learning performance, self-efficacy, and PDRLS use intention. Acharya and Sinha [53] in their work have used Direct Hashing and Pruning (DHP) algorithm to generate 'optimized' concept map of learning. A learning sequence was obtained from it for remediation purpose. A working prototype of the proposed model was implemented using Android emulator. The effectiveness of the proposed method was studied using multivariate regression analysis. Various concept maps were generated on varying the hash function. It was found that modulo division hash function generated the most optimized concept map [54].

Compared to E-Learning environment the number of personalized applications of concept maps in M-Learning environment has been few and far between. As hinted earlier, due to limited computing power automated generation of concept maps here is distant reality. Thus most of the learning systems developed in mobile environment have stored concept maps in the form of database tables. Concept maps mostly have been used here for organizing and structuring knowledge. A ubiquitous learning website has been developed by Chen et al. [55] to provide learning guidance to students. The learning system can be accessed by PCs' as well as mobile devices. They have assumed a concept map for an object oriented programming



course. A brief summary of some of these learning systems are shown in Table 2.

**5.2. Concept Maps in Collaborative Applications**

Following Jong’s [47] classification on the basis of applications, it has been found that concept mapping technique has mostly been used for collaborative applications. In typical applications, learners have been found to develop their own concept maps based on domain knowledge. This concept map is then finalized after discussion with peers. Support may also be taken from the instructors or

materials from the web. Some of these applications in E-learning and M-Learning environment are next discussed.

Over the years several researchers have used collaborative learning strategy in E-Learning environment. Chang et al. [56] has proposed a four stage web based collaborative inquiry learning system that uses the web as source of knowledge. Concept map has been used for organizing this knowledge. These stages of learning were used to generate the concept map by a set of 17 students in a university in Taiwan.

**Table 2.** Summary of some diagnostic concept maps developed in personalized environment

Reference	Data Mining heuristic used	Concept map objective	Entities on which data mining heuristics are applied	Entities involved in learning
Hwang [33, 48]	None	Diagnosing student misconceptions	Student test results, concept weights	Learner
Jong et al. [47]	Sequential Probability Ratio Test	Identify missing concepts and faults in a concept map	Test item database, teaching material, concept map drawn by a specialist	Learner, concept map specialist
Chen et al. [50]	Apriori	Diagnosing common learner misconceptions	Test item and course database	Learner
Tseng et al. [49]	Fuzzy set theory	Adaptive learning guidance for learners	Learner’s historical test records, test item-concept mapping database	Learners, educational experts
Chen et al. [52]	None	Student’s learning sequence for learning status	None	Learners, teachers
Lee et al. [34]	Apriori	Diagnose learning barrier and misconceptions	Test item-weight relation, learners test portfolio	Learner, teacher
Chen [65]	Pathfinder network analysis	Improved learning performance, self efficacy and use intention	Courseware, learner account, conceptual relationship and user profile databases	Learner
Acharya and Sinha [53, 54]	Direct Hashing and Pruning	Determining optimal learning sequence	Student test results, concept weights	Learner

A 17 item 4 point Likert scale was used to find the opinion of students in web based learning. Simone et al. [57] has used three tools namely

student collaboration, concept mapping and electronic technologies to foster knowledge growth among university students. 26 students

were divided into groups of 3 to 5 students each and were asked to generate concept maps on a given topic in learning theory. They found that students preferred sharing concept maps between themselves and is a good tool for generating and structuring ideas. Stayanova[58] have conducted an experimental study to investigate the learning effectiveness of concept map in Computer Supported Collaborative Problem Solving Design (CSCPSD). Three scenarios of interaction were investigated: distributed, moderated and shared. It was found that shared scenario was most appropriate for establishing a supportive environment for CSCPSD.

In mobile environment collaborative learning strategy has generally been applied to learning those subjects that have two learning components: a theoretical component taught in a class room and a practical component learnt using field work. These groups collaborate using tools like SMS, e-mail etc. Hwang et al. [35] has proposed a mobile learning approach based on concept maps with remediation mechanism. Two concept maps were constructed. The initial one was constructed by the teacher. Students then developed their concept maps based on the knowledge learnt from the text books in the computer class room. These concept maps were then compared. The concept map was finalized when students performed relevant field work in the subject. Lai et al. [59] reports their implementation of a handheld concept mapping tool to support cooperative learning in a nursing class. They used PicoMap software to enable students construct concept maps. They found that handheld tools enhanced interaction among students when aided by proper class management and technology support. Silander et al. [60] has used a Mobile Collaborative Concept Mapping (MoCoCoMa) learning system using the SMS property of mobile phones. The application software consists of PHP and

Java Applet module and the database shared with these. The system was tested for a set of students for studying natural science courses. Acharya and Sinha [61] in their work developed a collaboration schema in which a set of students studied a subject having both theory and practical components by exchanging SMS. A prototype learning system was again developed using Android emulator. A class of students learning the course "Introduction to Java Programming" were divided into groups and asked to study using the proposed method for evaluating their group dynamics.

A comparative study of some of these learning systems are presented in Table 3 on the basis of the collaboration objective, environment, application and the learning theory modeled in collaborative applications.

### 6 Summary of EDM Algorithmic Aspects

The previous sections discussed the web based learning systems developed in personalized and collaborative environment using EDM techniques with concept maps. This section attempts to summarize the general features of these algorithms used in personalized and collaborative learning applications discussed above and propose a set of parameters for comparing them.

The most important aspect of concept map generation using EDM techniques is the *computational complexity of the map construction algorithm*. It has been observed that for P-Learning general data mining algorithms like Apriori [34,50], DHP [33,54], SPRT [47] have been used for concept map computation. For C-Learning concept maps have been constructed based on discussion between the learners [57,59,61] and thus EDM techniques have not been used. However in both the types of environments the most important factor in complexity computation is the number of learners ( $n$ ) involved in concept map generation.

**Table 3.** Summary of some concept mapping techniques used in collaborative environment

Reference	Collaboration Objective	Collaboration Environment	Learning theory modeled	Collaborative Application
Simone [57]	Find effectiveness of group based electronic concept maps in identifying and organizing course contents	Concept mapping software 'Inspiration'	Constructivism	Construction of concept maps on 'Learning theories' by several groups and sharing these
Chang et al. [59]	Construction of collaborative inquiry learning model	Inquiry and discussion using the web	Inquiry learning	Developing concept maps on 'Pyramid Construction' individually and revising it on the basis of knowledge learnt from web and peers
Silander [60]	Combine class room activities with simultaneous field exploration	Mobile Collaborative Concept mapping (MoCoCoMa)	Progressive Inquiry	Two groups of students, one in class room and another in forest collaborate to construct concept maps of tree species.
Hwang et al. [35]	Construction of interactive concept maps oriented mobile learning system	Interactive concept maps-Oriented Mindtool for M-Learning (ICM <sup>3</sup> )	Not reported	Constructing concept maps based on class room knowledge and reviewing in field work using ICM <sup>3</sup> .
Acharya and Sinha[61]	Learning by concept map construction by exchanging SMS	Collaborative Learning System (CLS) developed using Android emulator	'Extended' Theory of Meaningful Learning	Two groups of students, one learning theory and another practical collaborate to develop association rules which form the edges of the concept map

Equally important is the *inputs required for construction of concept map*. Concept maps are generally constructed on the basis of certain prior information. This could be student's historical test records [33,48], learners perception of the subject or learning inputs received from the peers or the web [59,60]. Student's historical test records have traditionally played an important role in concept map computation in P-Learning environment. Sometimes instructor inputs [34] have also been used. EDM techniques have been applied to these for deriving

concept maps. On the other hand, student's knowledge of the subject has often been used in C-Learning environment as input to EDM techniques [57,59].

Web based learning systems have often used various *learning theories* based on which they have been implemented. These learning theories propose a set of step based on which learning is conducted. These steps have been implemented in the form of modules while designing the system architectures. Typical examples of learning theories used for learning system construction are Item

Analysis for Norm Referencing [52], Activity Attention Network [49] and Theory of Meaningful Learning [53]. In this context no difference was observed between P-Learning and C-Learning environment.

Next in line is *computation of confidence between concepts*. Confidences are used in determining the degree of association between the concepts. Acharya and Sinha [62] have used them for proposing weighted concept maps. In P-Learning environment they are typically computed using the algorithms used for determining association rules between the concepts. In C-Learning environment however confidences may be assigned by the learner intuitively. Weights may also be assigned to the concepts to indicate their degree of importance [35,61].

Just the computation of concept map may not be enough for all types of applications. Optimization of the map may be required in certain circumstances. One such optimization is *removal of cycles in concept maps*. A concept map containing several concepts is almost found to contain cycles. These cycles introduce a certain degree of redundancy. If a cycle exists between two concepts, it is impossible to determine which concept should be learnt first. Various researchers have proposed variety of graph algorithms for removal of redundancy. In P-Learning the general consensus is removal of the edges with least weight which removes cycles [34,49,54]. C-Learners however may avoid cycles based on observation [59,61]. Even after this if cycles exist, algorithms similar to P-Learning environments may be used.

The usage of EDM techniques is again dependent on the *focus area of the developed application*. Data mining algorithms are often used in P-Learning environments whose major application area is remedial learning [33,34,47]. C-Learners however rely more on discussion among themselves [60,61]. However, for both types of learning environments, web based tools have been used for prototype development. As examples Tseng et al. [49] have used PHP and MySQL for a P-Learning application whereas Android

emulator has been used by Acharya and Sinha [53,61] for both types of applications.

Finally the *efficiency of the learning system is determined in terms of learning result*. For both P-Learning and C-Learning environments the efficiency is determined by conducting tests before and after the learning process and applying certain statistical algorithms on these test results. Statistical algorithms include t-tests, ANOVA and multivariate regression analysis [33,35,50,52,65]. A survey is also used to evaluate learner satisfaction. Additionally C-Learning systems often use these statistical methods to determine the collaboration dynamics [61].

Based on the literature surveyed in the previous section, personalized and collaborative web learning systems developed using concept maps are compared parametrically in Table 4. A serial # is attached to each parameter for ready reference. On observation it is found that for P-Learning serial #s 1, 3, 4, 5, 7 use EDM techniques extensively whereas for C-Learning only serial #s 3, 7 use EDM techniques. Thus it can be concluded without a shadow of doubt that P-Learning environment uses EDM techniques a lot more compared to C-Learning environment.

## 7 Limitations of these Studies

The features of P-Learning and C-Learning applications pointed out in the previous sections are not without limitations. This section points out the limitations of these studies so that researchers in future may address them. Firstly, most of the researchers implemented the modules for P-Learning and C-Learning systems on the basis of certain established learning theories. As an example, Tseng et al. [49] in their study have used Item Analysis for Norm Referencing for developing their learning system in P-Learning environment. Similarly Acharya and Sinha [53] have used the theory of meaningful learning for their P-Learning system. For generation of C-Learning system they have added a new phase to this system and hence called it 'extended' theory of meaningful

learning [61]. Learners and academic administrators however often fail to get a deeper picture of the relationship between learning theories and their proposed implementation using EDM techniques.

**Table 4.** EDM algorithmic aspects of concept maps computation in personalized and collaborative applications

Serial #	EDM algorithmic aspect	Personalized Application	Collaborative Applications
1.	Computational complexity for concept map construction (n-number of learners)	For m test-items $O(mn)$ [34]	$O(n)$ for heterogeneous group creation; $O(nt)$ for message exchange (t is the number of messages exchanged per learner) [61]
2.	Inputs for constructing concept maps	Learner's historical test records [33]	Learner's subject knowledge [58]
3.	Use of learning theory	Extensive use of learning theory for generating architectural modules [49,52,53]	Similar procedure used [61]
4.	Computation of confidence between concepts	Computed automatically from association rules using Bayes theorem [34,53]	Intuitively assigned by the learners [35,61]
5.	Cycle removal algorithm	Removal of edges with least weights which creates cycles [34,49,53]	Cycles avoided by observation [59,61]
6.	Focus Area of the developed application	Diagnostic and remedial Learning [33,34,49,50,65]	Learning a subject using discussion[57,59,60]
7.	Efficiency of the algorithms in terms of learning result	Regression analysis to estimate the degree of significance between pretest and post test marks; surveys used to estimate learner satisfaction [33,35,50,52,65]	Similar procedure used [35,37,59, 60]. Statistical algorithms used in evaluating collaboration dynamics [61]

Secondly, as indicated in the introductory section, KDD process contains multiple steps like data cleaning, integration, selection, transformation, mining, pattern evaluation and knowledge presentation [2]. However in almost all the applications discussed above, the focus is on using EDM techniques for learning system construction only. Web learning applications are to be developed which integrates various applications like preprocessing student data, learning objects development and storage, designing test

modules along with data mining functionalities.

Thirdly, the domain of use of P-Learning and C-Learning has not yet been clearly established. In general terms, when certain remediation measure is to be designed for a small group of learners P-Learning has been used. Typical examples are Lee et al. [34], Chen et al. [50], Jong et al. [47]. C-Learning has been used when a group of learners wanted to study together. Typical examples are Simone [57], Chang et al. [59], Silander [60] etc. Acharya and Sinha[61] have used C-

Learning for two groups of learners learning a subject simultaneously, one group learning in theory class and another group learning in practical class. This classification is however too vague to have any conclusive effect.

Finally, using data mining algorithms in constructing web based learning systems is not an easy task. Proper understanding of functions like classification, clustering, association along with a host of statistical and machine learning algorithms is necessary. This process may be cumbersome and time consuming for most learners, educators and academic administrators.

## 8 Conclusion

This article explores the EDM aspects of concept map construction in web based learning environment for personalized and collaborative learning. For this purpose a brief overview of the applications of EDM in this task domain is first presented. The applicability of web based learning in this approach is next discussed. Finally, web based applications using EDM approaches often require a tool for knowledge organization. This article explores the use of concept map for this purpose. Finally a detailed survey of the pioneering web based learning systems developed using EDM and concept mapping techniques are presented and their algorithmic aspects compared. Limitations of the current research have also been pointed out.

Data mining algorithms have been widely used in construction of concept maps in web based learning environments. There are several promising directions to extend the work already done by the researchers. Firstly, recommender systems may be used that will help students identify learning patterns before appearing for a test. As an example, before appearing for a test on concept X, a student may like to know the concepts learned by his predecessors who had appeared for a test on that concept. Similar methods may be employed to find the learning objects accessed by students who have performed well in the exams. A recommender system using concept maps and collaborative tagging has been proposed by Kardan et al. [63]. The proposed

method performs a mapping between tags and the concepts learnt by the student for identifying his learning deficiencies. However collaborative filtering methods suffer from disadvantages like requiring huge data sets and high computational power to make correct recommendations [64]. In this direction the use of content based, knowledge based and hybrid recommender systems may be investigated [65].

Graph based applications have often been used to enhance EDM and concept mapping techniques. Acharya and Sinha [62] proposed the use of weighted concept maps to compute the learning path that has highest degree of importance from learner point of view. EDM techniques may be devised to find the cluster of students using similar learning patterns, i.e. the set of students who use the same learning paths. Certain data mining algorithms may then be applied on this student cluster data set to find interesting patterns within themselves. Several researchers [34,49,53] proposed removal of cycles in the concept map by deleting edges with minimal weight which creates cycles. However this strategy may not be efficient if the number of concepts is large. Another application could be to detect cycles in the concept map while they are generated in an automated fashion.

Finally, the association rules and the weights between the concepts for personalized applications were generated in various studies [34,53] by applying varied data mining algorithms on historical test records of the students. This concept map is thus dependent on the particular data mining algorithm applied on the test records. Thus a generic method could be developed that would generate a unique concept map for a particular test record.

## REFERENCES

- [1] U. M. Fayyad, G. Piatetsky-Shapiro, P. Smyth, & R. Uthuramy, *Advances in Knowledge Discovery and Data Mining*, AAAI/MIT Press, 1996.
- [2] J. Han, M. Kamber, & J. Pei, *Data mining: concepts and techniques*. Elsevier, 2011.
- [3] R. O. Duda, P.E. Hart & D. G. Stork,

- Pattern Classification. 2nd Edition, John Wiley & Sons, Inc., New York NY, 2001.
- [4] A. K. Jain, R. C. Dubes, Algorithms for Clustering Data. Englewood Cliffs, NJ: Prentice-Hall, 1998.
- [5] C. Romero, & S. Ventura, "Educational data mining: A survey from 1995 to 2005", In: Expert systems with Applications, 33(1), pp. 135-146, 2007.
- [6] C. Romero, & S. Ventura, "Educational data mining: a review of the state of the art", In: IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews, 40(6), pp. 601-618, 2010.
- [7] S. Fiedler, & T. Väljataga, "Personal learning environments: concept or technology?", International Journal of Virtual and Personal Learning Environments, 2(4), pp. 1-11, 2010.
- [8] E. Pogorskiy, "Using personalization to improve the effectiveness of global educational projects", In: E-Learning and Digital Media, 12 (1): pp. 57-67, 2010.
- [9] M. Fullan, Michael Fullan's Answer to "What is Personalized Learning?", Microsoft Partner Network, 2009.
- [10] P. Dillenbourg, What do you mean by collaborative learning?. In: P. Dillenbourg (Ed) Collaborative-learning: Cognitive and Computational Approaches. pp.1-19. Oxford: Elsevier, 1999.
- [11] K. Bruffee, Collaborative Learning. The Johns Hopkins University Press. Baltimore, pp. 28-51, 1993.
- [12] A. Ingram, Using web server logs in evaluating instructional web sites, In: Journal of Educational Technology Systems, 28(2), pp. 137-157, 1999.
- [13] G. J. Hwang, P. S. Tsai, C. C. Tsai, J. C. R. Tseng, "A novel approach for assisting teachers in analyzing student web-searching behaviors", In: Computer & Education, 51(3), pp. 926-938, 2009.
- [14] J. Gibbs, M. Rice, M., "Evaluating student behavior on instructional web sites using web server logs", In: Ninth Sloan International Conference on Online Learning, Orlando, pp. 1-3, 2003.
- [15] A. Anaya, & J. Boticario, "A data mining approach to reveal representative collaboration indicators in open collaboration frameworks", In: International Conference on Educational Data Mining, Cordoba, Spain, pp. 210-218, 2009
- [16] E. Heathcote, S. Dawson, S., "Data Mining for Evaluation, Benchmarking and Reflective Practice in a LMS", In: World conference on E-learning in corporate, government, healthcare & higher education, Vancouver, Canada, pp. 326-33, 2005..
- [17] C. Pahl, & C. Donnellan, "Data mining technology for the evaluation of web-based teaching and learning systems", In: E-learning Congress, Montreal, Canada, pp.1-7, 2003.
- [18] A. Merceron & K. Yacef, "Interestingness Measures for Association Rules in Educational Data", In: International Conference on Educational Data Mining, Montreal, Canada, pp. 57-66, 2008
- [19] J. Mostow, J. Beck, H. Cen, A. Cuneo, E. Gouvea, & C. Heiner, "An educational data mining tool to browse tutor-student interactions: Time will tell!", In: Proceedings of the Workshop on Educational Data Mining. pp. 15-22, 2005
- [20] P. Yu, C. Own, & L. Lin, "On learning behavior analysis of web based interactive environment", in: International Conference on Computer and Electrical Engineering, Oslo/Bergen, Norway, pp. 1-9, 2001.
- [21] C. J. Tsai, S. S. Tseng, & C. Y. Lin, "A Two-phase fuzzy mining and learning algorithm for adaptive learning environment" In: International Conference on Computational Science, San Francisco, pp. 429-438, 2001
- [22] Z. Ibrahim, & D. Rusli, "Predicting students' academic performance: comparing artificial neural network, decision tree and linear regression", In: Annual SAS Malaysia Forum, Kuala Lumpur, pp. 1-6, 2007.
- [23] A. Arnold, B. Scheines, R. Beck & J. E. Jerome, "Time and Attention: Students,

- Sessions, and Tasks”. In: AAAI2005 Workshop on Educational Data Mining, Pittsburgh, pp. 62-66, 2005.
- [24] R. Stevens, A. Giordani, M. Cooper, A. Soller, I. Gerosa, & C. Cox, “Developing a Framework for Integrating Prior Problem Solving and Knowledge Sharing Histories of a Group to Predict Future Group Performance” , In: International Conference on Collaborative Computing: Networking, Applications and Worksharing, Boston, pp. 1-9, 2005.
- [25] F. H. Wang, & H. M. Shao, “Effective personalized recommendation based on time-framed navigation clustering and association mining” *Expert Systems with Application*, 27, pp. 365-377, 2004.
- [26] T. Y. Tang, & G. Mccalla, “Student modeling for a web-based learning environment: a data mining approach”, In: Conference on Artificial intelligence, Edmonton, Canada, pp. 967-968, 2002.
- [27] D. Perera, J. Kay, I. Koprinska, K. Yacef, & O. R. Zaïane, Clustering and Sequential Pattern Mining of Online Collaborative Learning Data. In: *IEEE Transaction on Knowledge and Data Engineering*, 21(6), pp. 759-772, 2009.
- [28] D. Zakrzewska, “Cluster analysis for user’s modeling in intelligent e-learning systems”, In: International Conference on Industrial, Engineering & Other Applications of Applied Intelligent Systems, Poland, pp. 209-214, 2008.
- [29] R. Nugent, E. Ayers, & N. Dean,. Conditional subspace clustering of skill mastery: identifying skills that separate students. In: International Conference on Educational Data Mining, Cordoba, Spain, pp. 101-110, 2009.
- [30] C. Tang, R. Lau, Q. Li., H. Yin, “Personalized courseware construction based on web data mining”. In: First International Conference on Web Information Systems Engineering, Hong Kong, China, pp. 204-211, 2000.
- [31] C. C. Kiu, C.S. Lee,., “Learning Objects Reusability and Retrieval through Ontological Sharing: A Hybrid Unsupervised Data Mining Approach”. In: IEEE Conference on Advanced Learning Technologies, Japan, pp. 548-550, 2007.
- [32] T. Orzechowski, S. Ernst, & A. Dziech, “Profiled search methods for e-learning systems”, In: International Workshop on Learning Object Discovery & Exchange at European Conference on Technology Enhanced Learning, Crete, Greece, pp. 1-10, 2007.
- [33] G. J. Hwang, “A conceptual map model for developing intelligent tutoring systems”. In: *Computers & Education*, 40(3), 217-235, 2003.
- [34] C. H. Lee, G. G. Lee, & Y. Leu, Y, “Application of automatically constructed concept map of learning to conceptual diagnosis of e-learning”, In: *Expert Systems with Applications*, 36(2), pp. 1675-1684, 2009.
- [35] G. J. Hwang, P. Wu, & H. Ke, “An interactive concept map approach to supporting mobile learning activities for natural science courses”, In: *Computers & Education*, 57(4), pp. 2272-2280, 2011
- [36] J. L. Moore, C. Dickson-Deane, & K. Galyen, “E-Learning, online learning, and distance learning environments: Are they the same?”, In: *The Internet and Higher Education*, 14(2), pp. 129-135, 2011.
- [37] E. Cherian, & P. Williams, “Mobile learning: The beginning of the end of classroom learning”, In: *Proceedings of the World Congress on Engineering and Computer Science*, pp. 22-26, 2008.
- [38] T. Georgiev, E. Georgieva, & A. Smrikarov, “M-learning: A New Stage of E-Learning”. In: International Conference on Computer Systems and Technologies-CompSysTech, Vol. 4, No. 28, pp. 1-4, 2006.
- [39] J. Novak, & A. Cañas, “The theory underlying concept maps and how to construct and use them”, In: *Reflecting Education*, 3(1), pp. 29-42, 2008.
- [40] J. Anhalt, A.Smailagic, D. Siewiorek, F. Gemperle, “Toward context-aware computing: experiences and lessons”, In: *IEEE Intelligent Systems*, 6(3), pp. 38-46, 2007



- [41] W. B. Johnson, L. Neste, & P. C. Duncan, "An authoring environment for intelligent tutoring systems". In: Proceedings IEEE International Conference on Systems, Man and Cybernetics, pp. 761-765, 1989.
- [42] V. Vasandani, T Govindaraj, & C. Mitchell, "An intelligent tutor for diagnostic problem solving in complex dynamic systems". In: Conference Proceedings Systems, Man and Cybernetics, IEEE, pp. 772-777, 1989.
- [43] M. Virvou, & E. Alepis, "Mobile educational features in authoring tools for personalised tutoring", *Computers & Education*, 44(1), pp. 53-68, 2005.
- [44] S. A. Kazi, "VocaTest: An intelligent tutoring system for vocabulary learning using the M-Learning approach".
- [45] J. Psocka, S. A. Mutter, *Intelligent Tutoring Systems: Lessons Learned*, Lawrence Erlbaum Associates, 1998
- [46] D. Ausubel, *The Psychology of Meaningful Verbal Learning*. New York: Grune & Stratton, 1963
- [47] B. Jong, T. Lin, T., Y. Wu, & T. Chan, "Diagnostic and remedial learning strategy based on conceptual graphs", In: *Journal of Computer Assisted Learning*, 20(5), pp. 377-386, 2004.
- [48] G. J. Hwang, C. L. Hsiao, & J. Tseng, "A computer-assisted approach to diagnosing student learning problems in science courses", In: *Journal of Information Science and Engineering.*, 19(2), 229-248, 2003.
- [49] S. S. Tseng, P. C. Sue, J. M. Su, J. M. Weng, & W. N. Tsai, "A new approach for constructing the concept map", In: *Computers & Education*, 49(3), pp. 691-707, 2007
- [50] C. M. Chen, Y. L. Hsieh & S. H. Hsu, "Mining learner profile utilizing association rule for web-based learning diagnosis", in: *Expert Systems with Applications*, 33(1), pp. 6-22, 2007.
- [51] M. Hagiwara, "Self-organizing concept maps", In: *Proceedings of the IEEE International Conference on Systems, Man and Cybernetics*, Vancouver, IEEE, pp. 447-451, 1995.
- [52] L. H. Chen, "Enhancement of student learning performance using personalized diagnosis and remedial learning system", *Computers & Education*, 56(1), pp. 289-299, 2011.
- [53] A. Acharya & D. Sinha, "An Intelligent Web based system for Diagnosing Student Learning Problems using Concept Maps", In: *Journal of Educational Computing Research*, SAGE Publishers.
- [54] A. Acharya & D. Sinha, "Smart Concept Map Generation for Learning Using Class of Hash Functions", *Smart Computing Review*, 5(4), pp. 329-339, 2017.
- [55] G. D. Chen, C. K. Chang & C. Y. Wang, "Ubiquitous learning website: Scaffold learners by mobile devices with information-aware techniques", *Computers & Education*, 50(1), pp. 77-90, 2008
- [56] K. E. Chang, Y.T. Sung & C.L. Lee, "Web-based collaborative inquiry learning", *Journal of computer assisted learning*, 19(1), pp. 56-69, 2003
- [57] C. D. Simone. R. F. Schmid & L. A. McEwen, "Supporting the Learning Process with Collaborative Concept Mapping Using Computer-Based Communication Tools and Processes", *Educational Research and Evaluation*, 7(2), pp. 263-283, 2001.
- [58] N. Stoyanova, & P. Kommers, "Learning effectiveness of concept mapping in a computer supported collaborative problem solving design", In: *First European International Conference on Computer-Supported Collaborative Learning*, 2001.
- [59] C. Lai, C. Wu, H. Kao, S. Chen, "Using a Handheld Concept Mapping Tool for Cooperative Learning", In: *International Journal of Artificial Intelligence*, 6(2), pp. 77-86, 2007
- [60] P. Silander, E. Sutinen & J. Tarhio, "Mobile Collaborative Concept Mapping – Combining Classroom Activity with Simultaneous Field Exploration", In: *Proceedings of the 2nd IEEE International Workshop on Wireless and Mobile Technologies in Education*, 2004.

- [61] A. Acharya & D. Sinha, A Web based Collaborative Learning System using Concept Maps: Architecture and Evaluation, Handbook of Research on Natural Computing for Optimization Problems, IGI Publishers, 2016.
- [62] A. Acharya & D. Sinha, "A weighted concept map approach to generate learning guidance in science courses",. In: Proceedings of the 3rd International Conference on Information Systems Design and Intelligent Applications (pp. 143-152). Springer International Publishing, 2015
- [63] K. A. Kardan, S. Abbaspour, & F. Hendijanifard, "A hybrid recommender system for e-learning environments based on concept maps and collaborative tagging". In: Proceedings of the 4th International Conference on Virtual Learning, Vol. 156, pp. 300-307, 2009.
- [64] S. Lee, J. Yang, & S. Y.Park, "Discovery of hidden similarity on collaborative filtering to overcome sparsity problem", In: Discovery Science, Springer Berlin Heidelberg, pp. 396-402, 2004
- [65] A. Kobsa, "Generic user modeling systems",. In: The adaptive web, Springer Berlin Heidelberg, pp. 136-154, 2007.



**Anal ACHARYA** is Assistant Professor in the Department of Computer Science in St Xavier's College, Kolkata. He was the Head of the Department of Computer Science from the period of 01.01.2009 to 30.06.2013. His present research interest is Educational Data Mining (EDM). He has about 20 years of experience in undergraduate and post graduate teaching & supervised several post graduate dissertations. He has several accepted papers in International Conferences and Journals. He has been awarded Ph.D. (Tech.) in Computer

Science and Engineering by University of Calcutta under the supervision of Prof. Devadatta Sinha.



**Devadatta SINHA** is currently Professor in the Computer Science and Engineering Department in University of Calcutta, Kolkata. He was the Head of the department on several occasions. His research interests include Distributed Systems. He has over 30 years of experience and has supervised several Ph.D. students.