

Online Contentbase Learner System

Supaporn Bundasak
Department of Computer Science
Kasetsart University, Thailand.
e-mail:supaporn@src.ku.ac.th

Anongnart Srivihok
Department of Computer Science
Kasetsart University, Thailand.
e-mail:anongnart.s@ku.ac.th

Abstract

At present, there are many online learning systems which are available and provided on the Internet. Almost all of these systems are static contents and test banks. To improve the teaching and learning on the Internet, this study proposes a new learning system that can be adjusted to the knowledge and understanding of students. This system uses K-Means Algorithm to cluster the student by pretest scores. Then, the lessons are provided for each appropriate clusters. The new learning system is developed to provide lessons according to students knowledge and performance. This will enhance self leaning on the Internet and it will save time for learning in the future.

Key words: *Internet, online learning, knowledge, contentbase, K-Mean Algorithm, cluster, score.*

1 Introduction

Normally the E-learning curriculum is the static module which could not be able to be adjusted to the learner abilities and qualities. It is difficult to identify the level of the learner ability after they have learned through the E-learning course. E-learning course is designed for self-study when learners have any questions in the lesson, they can go back to the former lesson which is the same content and procedures of learning and teaching method. They can

go back to the former lesson until they feel comfortable to go to the next lesson.

At the present, E-learning course has been developed to support the interactive online program also it supports the student for self-study. By advancing and developing the internet learning meets its efficiency, therefore the programs are designed to be intelligent and supportive for the different level of learners by using algorithm of machine learning such as K-Mean Clustering. The K-Mean Clustering Algorithm will classify the learners into groups on the basis of their background, then students are assigned with the right and proper contents. The reason to select K-Mean Clustering algorithm is that the data has been divided obviously and it is able to analyse quickly, precisely and validly. Also it is not too complicated to a web program design.

This study aims to use K-Mean Clustering Algorithm for clustering students in the E-learning course. This will bring the learners with the same level together. Learning in the right and proper curriculum, the learner will learn the customised knowledge and system has ability to supply information efficiency. Moreover, this study will develop the internet E-learning supporting to future learning and teaching.

2 Literature Review

Apart from the related research the study will find out the method using in online learning and teaching research. At the present the document examination was found that the learning and teaching modules have very normal technique test for the learners after the lesson. There has been an interesting method by developing a web base followed the Data Mining (Zaiane 1999). The procedures are as followed:

- (1) Eradicate unrelated data.
- (2) Classify the phenomenal session.
- (3) Category in locked file appearing in learning activities.
- (4) Consider the completed route.
- (5) Group the learner sessions and bring to classify the learning sessions.
- (6) Gather the other data related to the learners and group the learners.
- (7) Learning and teaching having a conclusion of behavior of locked files learning which method is suitable for conclusion analysis to seek the method and the target of learning; however, it's not suitable for automatic interactive learning because the study found learner behaviors after they have finished the lesson.

The Online interactive learning (Cheng 2000) did not use any Artificial Intelligent Algorithm in the system. So the researcher was interested in using this algorithm to study the learner ability adapting by using Adaptive Test Engine or On-line Interactive and Adaptive Learning.

Josemh et al. (2000) had studied the Intelligent tutoring systems (ITSs) which comprises 5 parts as followed:

- (1) Student Model : it collects the main learner data since the beginning. This data is an input for Pedagogical Model.
- (2) Pedagogical Model : it is a teaching preparation process; newly topic

presented under this model and referring to the Student Model which is an input data which the teacher will give the response to the students by necessity.

(3) Domain knowledge : this section has the knowledge resources for a teacher using in teaching.

(4) Communication Model : this is a section using in communicating between teacher and the learner, and the layout on the screen.

(5) Expert Model : this section is like the Domain Knowledge but this model will record the students and the question answered by the teacher.

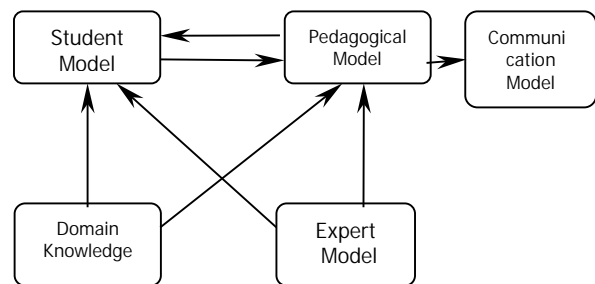


Fig 1. The illustration of intelligent tutoring system relationship. (Josemh et al. 2000)

Westhed (1999) had studied the related model. On-line interactive Modules telling the learner environments affected to learning efficiency performance that would be different when it's designed.

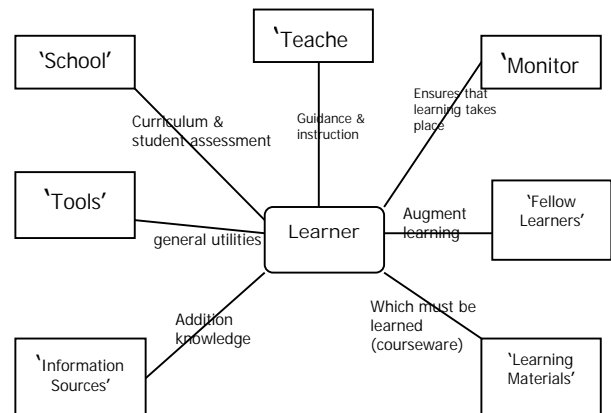


Fig 2. The illustration of On-line Course Modules. (Westhed 1999)

From the Course Module on the internet will bring to be a virtual university which is needed the environment as module on the internet. From the study the module has known the effects when we design the model.

Out of this the study methodology of existed learning and teaching models the researcher has studied Intelligent Learning Assistant in Multimedia-base Interactive e-learning.

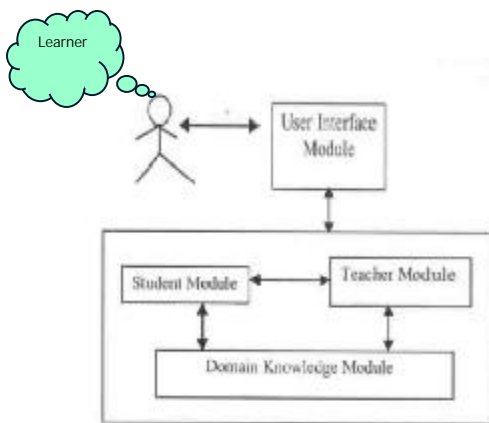


Fig. 3. The illustration of Architecture of Teacher assistant system.

The related research, Intelligent Tutoring by McFetridge and Heift. 1995 had used in German teaching needed by the education department. The model had originated from the needs to teach German and to have the teaching assistant. The error value of contingent feedback was set the student performances individually.

For the efficiency study performance the teacher was value by the teaching evaluation. Out of the student who had no experienced needed more details in teaching, the student who had an experiences needed the benefit of higher advanced explanation.

German Tutor has 4 components as followed. (1) Domain Knowledge (2) Analysis hierarchy (3) Student Module (4) Filtering Module.



Figure 1: The German Tutor

Fig. 4. The illustration of German Tutor (McFetridge and Heift. 1995).

From the former studies can be described German Tutor that there are two modules as the same; Domain Knowledge module ,Student Module but there has two as the differences.

- (1) Analysis hierarchy : Send the student learning data for statistic in Student Module.
- (2) Filtering Module : Filter the error value in student learning for learning repetition until they accomplish the lesson.

3 Methodology

3.1 Machine Learning

This method uses knowledge from analyzing the data sample to analyze and predict the outcome and then uses that outcome for the upcoming processes. Machine learning processes include clustering, classification and prediction.

-Clustering

Clustering involves dividing a set of data point non-overlapping grouping groups, or clusters, of points, where points in a cluster are “more similar” to one another than to points in other clusters. The term “more similar,” when applied to clustered points ,usually means closer by some measure of proximity. When a dataset is clustered ,every point is assigned to some cluster

,and every cluster can be characterized by a single reference point ,usually an average of the points in the cluster. Any particular division of all points in a dataset into clusters is called a partitioning(Vance Faber,1994).

-Classification

This process is to create an engine that can separate the data types into groups that has been set before. This process has 2 steps. The first step is educating the engine so that engine has knowledge to use in the next step which is separating the new coming data into related group.

-Prediction

This process uses the knowledge base from classification process to analyze and predict the interested subjects. From the study of On-line Interactive Learning (Cheng 1999) shows that the Artificial Intelligent Algorithm has not yet been used, so it makes analyst interested in apply on Artificial Intelligent Algorithm to On-line Interactive Learning. It will help developing the On-line learning program to adjust themselves to fit the knowledge abilities of the learners by using Adaptive Test Engine or On-line Interactive and Adaptive Learning.

3.2 K-Nearest Neighbor (Maleq et.al. 2002).

The study of Instance-Base Learning theory of the Machine Learning which developed from Nearest Neighbor. To make the Learning be able to stand more tolerant for data lost, we need to look at data set group at K data. To know the class of K by looking at the highest Class point, the point will be used to separate the data sample group that will be used for analysis. K value is used to predict the right answer by changing the attribute value. The positive thing about the Algorithm is

that it usually uses training data to create function that use to predict F(x). The answer is in Class follow an input value. Using the sample to create function. Paying attention to the group the close to the answer will predict the Class. By analyzing data that is nearby F(X) only, the f value located by K-Nearest Neighbor will be as follow:

$$f(x_q) \leftarrow \frac{\sum_{i=1}^k (x_i)}{k}$$

3.2 Clustering and the K-Means Algorithm.

Algorithm that has been used very often is the consolidate data by using Grouping tactic and Algorithm K-Means that developed specifically for using with Grouping a large data group. Grouping tactic will group data at the different point and the point in the data group will be the same. The grouping tactic that has been seen very often is classification of plant and animal by grouping them into specific group type such as species. The main propose of clustering with “Landsat data” type is to reducing data size and complication of the dataset Data reduction. It can be done by replacing the equal coordinates point.

The example of Grouping is the consolidation of the student marks that show in percent which divided to 5 cluster as follow A, B, C, D and F (Picture 5). The marks is the data points and reference point of each group is the average of the mark in that group. The letter that indicated grade can use as symbol to indicate the reference point instated of the number.



Fig 5. Clustering score data.

The data grouping can be described as follow. The cluster k will be set before the reference point Z_k in cluster l . It is a center point of cluster in case of one dimension data, for example the center point of the mark is the average mark in the cluster.

For multi dimension data which each data point contains many components. The center point of those components will be the same. Each component will be the average value of the same component of all data point in the cluster.

The two big problems of algorithm are (1) limitation in analysis large dataset and (2) the calculation method is not effective enough.

Each process has to calculate the distance between every data that analysed for cluster, even though algorithm will set the number of the group they want. The center point of the group that has been set might not be the real value of the data.

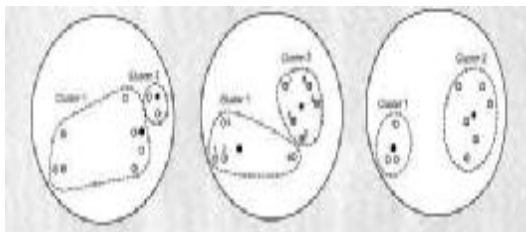


Fig 6. Clustering by using K-mean algorithm.

The conclusion of the K-mean algorithm process are as follow.

Iterative partitioning K-mean algorithm.

Step 1 Choosing the divider by starting at K clusters.

-partition - Beginning can be built by starting to create a set of k seed points which can be first k -patterns or

k -patterns that randomly choose from pattern matrix.

-partition – Different beginning can make different final clustering because the Algorithm is base on square error that can make it equal.

Step 2 Building new partition by setting each partition to center point of nearest cluster.

Step 3 Calculate the center of the cluster to be center point of the cluster.

-The set of K patterns that separated will be use by taking the center point of the data to be the first seed point and choosing the successive seed point that has the less distance from the seed point.

-square Error.

Step 4 Repeated in step 2 and step 3 until receive best value .

-partitions is updated by reassigning patterns go to clusters for trying to reduce the square error.

-Euclidean metric is use to calculate the distance between pattern and cluster center.

Step 5 The adjusting of cluster number by grouping and dividing cluster.

4 Design of the System

K-Mean Algorithm is selected in research because of the needs to arrange group score which identified the student ability into the three groups. From another algorithm for instance; Decision Tree which has a group of factor using in making a decision; however, it will bring only the result to analyze the test. An interesting algorithm is a neural network, giving a weight of input value which is suitable for not on-line data. That will give correct result because it repeated the data until having got the less error value but K-Mean Algorithm worked well on the on-line working. Because of data clustering or data filtering have not the supported software therefore many algorithms are

adapted properly for the clustering and filtering.

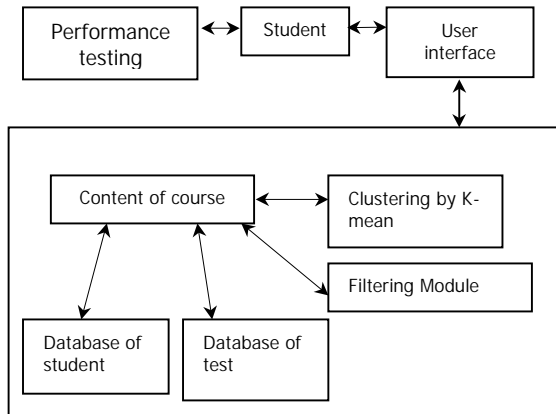


Fig 7. Research Model.

Explain the research model ,step of research will be as follow:

Prepare course materials for the study .-
 Select course of Electronic data Processing have Lesson plan.

Week 1 Introduce the computer. equipment

Week 2 MS-DOS

Week 3-4 MS-Windows

Week 5 Internet

Week 6-7 MS-Word

Week 8-9 MS-PowerPoint

Week 10-12 MS-Excel

The Lessen that will be use is MS-Word, and MS-Excel.

Table1. separates content by grouping student Performance in MS-Word.

less on	Content	Group of student Performance		
		low	medi um	high
1	Computer systems	ü		
2	Automatic office	ü		
3	Using computer with printing in office	ü	ü	ü
4	Introduction to printing	ü	ü	ü
5	Management of document file	ü	ü	ü
6	Create table in document file	ü	ü	ü

7	Put image and manage image	ü	ü	ü
8	Bring excel object, create Hyperlink and create mail merge			ü

Table2. Separate the content by grouping student performance in MS-Excel Lessons.

less on	content	Group of student Performance		
		low	medi um	high
1	Introduction to Excel	ü		
2	Typing and Edit data	ü		
3	Manage format and edit file	ü	ü	ü
4	Using calculate	ü	ü	ü
5	Create chart and put image in file	ü	ü	ü
6	Database management			ü
7	Apply excel for office			ü

-Database for processing group student will be as: (1) Question (2) Student (3) pretest scores and (4) final scores.

Designing the Master system and using the study of K-mean algorithm to develop the online learning and teaching, so it will be able to present themselves to fit the learner abilities. The K-mean algorithm is often used for grouping data, it is very reliable. Also the testing of algorithm has been done on the web page. Developing the algorithm and pulling the data to analyze through the algorithm.

-Information layout of monitor screen E-learning will be as: Lesson Plan, Course Objectives, Course Outline , Contents , and Test.

Developing and building the online learning program through the internet

by using the K-mean algorithm, the equipment that used to develop the system are : Pentium 4 RAM 256 Mb OS : Windows XP. Database QL server. Application : Editpus, Dreamweaver , Photoshop. Webserver : Internet Information Sever.

Experiment the effective of the program by grouping 2 groups of student, group 1 using online learning tactic and group 2 using the traditional class room study. Both groups have to do a test before they get in the experiment. The reason for doing that test is too mature the student abilities and knowledge. During the experiment the students in both group get in the system, they have to take MS word. The third and forth week study online and sixth, seventh week study in the classroom. They have to study MS Excel during the eighth and ninth week through online learning and the 10-12 week by classroom learning. At the end of the course, the student in both groups must take a rest again.

The effectiveness can be evaluated by compare the results of the tests before enter to the program and after using the program of the students in both group. The group that has a high percentage is the online learning group.

Analyze the experiment result and make the conclusion on the effectiveness of the algorithm. The result of the effectiveness of the algorithm turns out very well. It can be used to develop the system in the future.

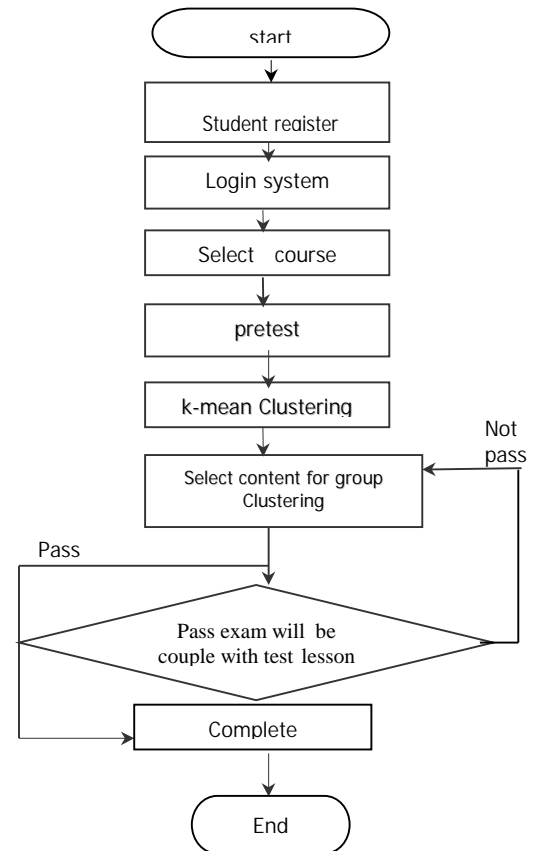


Fig 8. On-line learning procedure.

5 Results

Result of K-mean algorithm for clustering grouping learning by their knowledge ability.

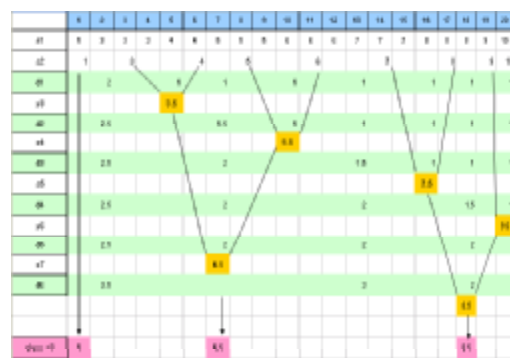


Fig 9. Data pretest in k-mean clustering algorithm.

Table3. Final Cluster Centers of test score for MS WORD lessons.

	Cluster		
	1	2	3
MS WORD	2.29	5.61	8.42

Table4. Final Cluster Centers of test score for MS EXCEL lesson.

	Cluster		
	1	2	3
MS EXCEL	1.65	4.37	7.38

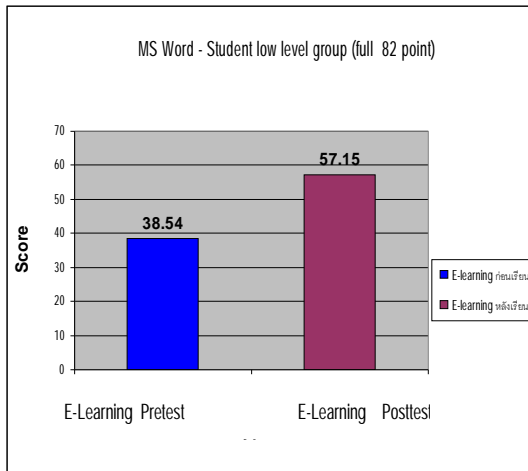


Fig 10. Score of student learn MS Word by e-learning pretest and posttest.

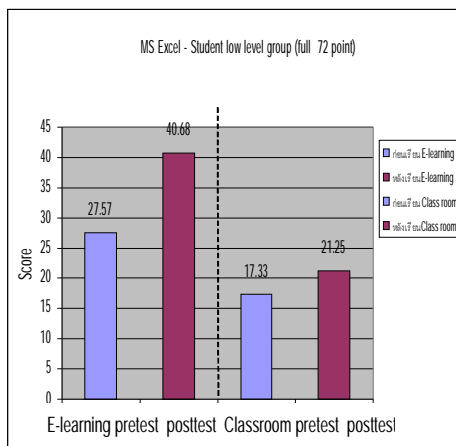


Fig 11. Score of student learn MS Excel by e-learning as :pretest and posttest , Classroom as pretest and posttest.



Fig 12. Login system.



Fig 13. Pretest exam system.



Fig14. Student report.

6 Conclusion

The approach for applying K-mean clustering algorithm for e-learning system has been introduced in this study. The students attended in this e-Learning system are clustered in three groups according to their pretest scores. This system provides contents based on group learning and their learning abilities. Finally was developed a prototype for an intelligent on-line learning by using an algorithm and database of learners to increase learner performance.

7 References

- [1] B. Joseph, S. Mia and H. Eirk, "Applications of AI in Education", 2000.
- [2] C.E. Beevers and D H Jackson, "Automatic Assessment-Mathwise and Beyond.department of Mathematics", Heriot-Watt University, ScotlandZ , 1994.
- [3] Cisco System , "Model of an E-Learning Solution Architecture for the Enterprise", 2001 .
- [4] F. Chang, H. Suarez and T. Cheng, "Web-Base Teaching Tools for On-line Interactive Learning", Polytechnic University, 1999 .
- [5] H. Pownall, "The CASTLE Project Creating On-line Assessment", Leicester University, 1997.
- [6] J. Cheng and R. Greiner, "Learning Bayesian Belief Network Classificatiors : Algorithms and System" . Department of computing Science . University of Alberta.Canada, 1999.
- [7] K. Maleq, D. Qin and P. William, "K-Nearest Neighbor Classification on Spatial Data Streams", Computer Science Department, North Dakota State University Fargo, USA, 2002.
- [8] M. Grobelnik, "Predicting content from hyperlinks" , department of Intelligent system J.Stefan Institute, 1997 .
- [9] M. Westhed, "Use of Web and Internet Technology in Teaching and Learning", Edinburgh Parallel Computing Center The University of Edinburgh, 1999.
- [10] R. Osmar, Zaiane, "Web Usage Mining for a Better Web-Based Learning Environment", Department of Computer Science University of Alberta Edmonton, Alberta Cannada , 2001.
- [11] S.L. Williamson and F. Lamb, "What can the Learning and Teaching Support Network offer you?", LTSN Engineering Loughborough University UK, 2001.
- [12] Timothy, K. Shih, "The specification and Implementation of a Virtual University Software System", Department of Computer Software University of Aizu. Japan, 2000.