2009 Ninth International Conference on Intelligent Systems Design and Applications

Developing Student Model using Kohonen Network in Adaptive Hypermedia Learning System

Bariah Yusob
Faculty of Computer System & Software Engineering,
Universiti Malaysia Pahang,
Lebuhraya Tun Razak, 26300
Kuantan, Pahang, Malaysia.
bariahyusob@ump.edu.my

Siti Mariyam Hj Shamsuddin
Soft Computing Research Group
Universiti Teknologi Malaysia,
81310 UTM Skudai, Johor,
Malaysia.
mariyam@utm.my

Nor Bahiah Ahmad
Soft Computing Research Group
Universiti Teknologi Malaysia,
81310 UTM Skudai, Johor,
Malaysia.
bahiah@utm.my

Abstract—**This paper presents a study on method to develop student model by identifying the students' characteristics in an adaptive hypermedia learning system. The study involves the use of student profiling techniques to identify the features that may be useful to help the researchers have a better understanding of the student in an adaptive learning environment. We propose a supervised Kohonen network with hexagonal lattice structure to classify the student into 3 categories: beginner, intermediate and advance to represent their knowledge level while using the learning system. An experiment is conducted to see the proposed Kohonen network's performances compared to the other types of Kohonen networks in term of learning algorithm and map structure. 10-fold cross validation method is used to validate the network performances. Results from the experiment shows that the proposed Kohonen network produces an average percentage of accuracy, 81.3889% in classifying the simulated data and 51.6129% when applied to the real student data.**

Keywords-Hypermedia Learning, Kohonen, Student Model

I. INTRODUCTION

Adaptive hypermedia learning system or AHLS is computer aided educational software that integrates two approaches which are the intelligent tutoring and educational based hypermedia technology. Intelligent tutoring focus on the knowledge and ability to adapt the learning sequences to match with the student's knowledge level. Educational based hypermedia technology on the other hand is a usage of media technology such as images, videos and voice recording in the learning material presentation including notes, exercises and more.

The aim of AHLS is to provide the learning material presentation to the individual student that has been adapted to the variety of backgrounds, knowledge level, learning behavior, objective and preferences. For that reason, the important issue in AHLS is to identify the student's unique features so that the best learning material can be provided to satisfy the student's need and the learning objectives can be successfully achieved.

One of the features in AHLS is the student is always been given a very huge and complex information and this let the student to become lost in hyperspace [1]. A few signs that the student is in this situation are: student fails to identity where he is; student can not get back to the previous page, student can not identify the page that contain useful information to him; and student can not identify important information regarding the learning material provided to him. This situation happen because he does not know which link he should go next to achieves his learning objectives.

Second problem will arise if there is a teacher whose task is to guide and supervise the learning activities through AHLS. It is very difficult for the teacher to control the students' activities and to identify their behavior while they are online [2].

Therefore, this research study the features that are available in the student model that can be learned, analyzed and accessed to identify the student's knowledge level in AHLS. Supervised Kohonen network using hexagonal lattice structure is developed and tested on the students' data to investigate its performances in learning the data and identify the students' knowledge level.

II. IDENTIFYING THE STUDENT MODEL

Student modeling is an analysis and prediction of student behavior and learning performances [3]. This process involves the development of data structure and reasoning mechanism to explore unique features for individual student and trace the student interaction in the AHLS environment. Example of the student features are behavior patterns, cognitive data such as knowledge, preferences and learning objectives and also non-cognitive data like emotional aspect and student personalities.

There were many examples of student features that have been introduced by the previous researchers. In this study, it is important to select the student model which contain features that influent directly to the student performance in the learning process. If not, process of identifying the student model will become complex and difficult. According to [4], among the features identified are the knowledge level, experience, backgrounds and student preferences. Apart from that, a few researchers suggested the use of age, gender, race, demographic data, preferences and others. However, the selection of the features must base on the types of knowledge that need to be gathered by individual cases.

978-0-7695-3872-3/09 $26.00 © 2009 IEEE
DOI 10.1109/ISDA.2009.103

Process of getting the student data is depends on the objective of the research. In this study, two student profiles are used to build the student model, which are the implicit and explicit techniques. Explicit data is information on the student that is easy and direct, which the system will simply ask the student what the information needed. This is done by using the questionnaire or registration form. Student has to fill in the personal and additional information in the registration form in order for them to start using the system. Usually, the questionnaire form is used to understand what the student wants from the system, such as preferences, preferable, necessity. Information gathered through this approach is static and sometimes doubtful because the system is unable to confirm the information provided by the student. There is a possibility that the student feel disturbed when asked to answer questions while using the system. Some students also may feel reluctant to give sincere answer because they worry it will invade their privacy.

Implicit approach is the alternative for the first approach that analyzes the student navigation behavior while using the system. Student usually do not realize that every movement and activities has been captured and recorded in a log file or can be stored in the system database. Through this technique, system can identify the student behavior without bringing distraction to the student. However, it also may cause the process of analyzing the student profile become complex and difficult to extract.

A. *Features in Student Model*

In this study, we integrate both student profile techniques to guarantee the effectiveness of the representation of the student as good as possible.

There are 4 attributes identified to represent the possible student features. These attributes are chosen based on the quality that has been suggested by [5] and [6]. We modify the features to adapt to the AHLS environment. The attributes are: learning time (x_1), number of backtracking (x_2), number of getting help (x_3) and assessment score (x_4). Specific information about the attributes is discussed as the following:

1) Learning time, x_1: Learning time is the time taken by a student to finish his learning activity in individual session. Based on [5], if the time taken by a student is more than the time estimated by the system, it shows that the student has problem in understanding the learning material provided to him. If the student is able to finish learning within the time estimation, he is considered as a fast learner and ready to proceed to the next part of learning.

When a student register to the AHLS to start the learning process, system will record the time the student spend on learning. Learning material is presented in a concept in the subject. The system will be given an estimated time for a student to finish learning in a single part. These parts are the types of learning material provided by the AHLS to satisfy the student preferences (theory, exercise, example, activity). Student can choose which part he prefers to use along the learning process. Learning time is calculated based on percentage of time taken to finish learning to the time estimated by the system. The estimated time is a time suggested by the expert in the subject learned by the student.

Definition of learning time, x_1:

Let subject taken is Data Structure, S,
There are several topics in this subject, T_i,
Therefore,

$T \subset S$, with $t_i = \{t_1, t_2, ..., t_m\} \in T$

In each topic (t_i), there are several sub-topic, $f_j = \{f_1, f_2, ..., f_n\}$ that consist of 4 defined learning material structure,

$$P = \{P_{note}, P_{example}, P_{exercise}, P_{activity}\},$$

with

$$P_{note} = \{P_{note\_1}, P_{note\_2}, ..., P_{note\_n}\},$$
$$P_{example} = \{P_{example\_1}, P_{example\_2}, ..., P_{example\_n}\},$$
$$P_{exercise} = \{P_{exercise\_1}, P_{exercise\_2}, ..., P_{exercise\_n}\},$$
$$P_{activity} = \{P_{activity\_1}, P_{activity\_2}, ..., P_{activity\_n}\}$$

For each topic, T,
Time estimated for each P is defined as:

$$p(t) = \{p_{note}(t), p_{example}(t), p_{exercise}(t), p_{activity}(t)\}$$

With

$$p_{note}(t) = \{p_{note\_1}(t), p_{note\_2}(t) ..., p_{note\_n}(t)\},$$
$$p_{example}(t) = \{p_{example\_1}(t), p_{example\_2}(t) ..., p_{example\_n}(t)\},$$
$$p_{exercise}(t) = \{p_{exercise\_1}(t), p_{exercise\_2}(t) ..., p_{exercise\_n}(t)\},$$
$$p_{activity}(t) = \{p_{activity\_1}(t), p_{activity\_2}(t) ..., p_{activity\_n}(t)\}.$$

Actual learning time for P,

$$p_{ac}(t) = \{p_{ac\_note}(t), p_{ac\_example}(t), p_{ac\_exercise}(t), p_{ac\_activity}(t)\}$$

With

$$p_{ac\_note}(t) = \{p_{ac\_note\_1}(t), p_{ac\_note\_2}(t) ..., p_{ac\_note\_n}(t)\},$$
$$p_{ac\_example}(t) = \{p_{ac\_example\_1}(t), p_{ac\_example\_2}(t) ..., p_{ac\_example\_n}(t)\}$$
$$p_{ac\_exercise}(t) = \{p_{ac\_exercise\_1}(t), p_{ac\_exercise\_2}(t) ..., p_{ac\_exercise\_n}(t)\},$$
$$p_{ac\_activity}(t) = \{p_{ac\_activity\_1}(t), p_{ac\_activity\_2}(t) ..., p_{ac\_activity\_n}(t)\}.$$

Therefore,

$$\text{Learning time, } x_1 = \frac{\sum p_{ac}(t)}{\sum p(t)} \times 100\% \qquad (1)$$

2) Number of backtrack, x_2 : Student navigation gives us information on student learning strategy and goal. [7] defines navigation as a method to solve a certain task. Internet users usually use a navigation strategies to find information. One possible way of identifying the navigation strategies is by counting how many internet user backtrack to the previous page [8].

[9] has identified 5 navigation strategies:

a) Scan: User scan the overall pages without go in through each of the pages.

b) Read: User open most of the pages and explore till he find pages that attract him.

c) Search: User open pages to find the information that he need only. He usually has objectives to achieve.

d) Explore: User open most of the pages and go in though all of them.

e) Wonder: Almost all pages are opened until it forms netted pages. This shows that the user's navigation has no structure and no objective..

[8] used this definition to identify the student navigation strategies through the number of backtracking. Number of backtracking shows that the student has not mastering the concept he has learned, lose direction or change his learning goal. Number of backtrack is counted by how many time the student reopen the previous page or note.

3) Number of Getting Help, x_3 : Help function is an important teaching tactic in AHLS to guide the student that has problem in understanding a concept or having problem in solving a question given by the system. With the help function, it can help refresh the student memory on previous facts [10].

In AHLS, help function is provided by listing the definition and explanation on specific term in the learning material. The more help given to the student, it shows that the student is having problem understanding the concept. Number of using help is counted on how many time the student click on help button during learning.

4) Assessment Score, x_4 : To access the student performance, an assessment is given at the end of learning session [5]. The assessment score is calculated based on percentage of correct answer.

B. Classificaton of Student Model

There are 3 student classes that have been defined to represent student knowledge level; advance, intermediate and beginner as shown in Table I:

TABLE I. STUDENT CLASSIFYING CRITERIA

Attribute	Beginner	Intermediate	Advance
Learning time, x_1	$x_1 > 100$	$80 \le x_1 \le 100$	$x_1 < 80$
Number of backtracking, x_3	$x_2 > 10$	$5 \le x_2 \le 10$	$x_2 < 5$
Number of getting help, x_3	$x_3 > 10$	$5 \le x_3 \le 10$	$x_3 < 5$
Assessment score, x_4	$x_4 < 60$	$60 \le x_4 \le 80$	$x_4 > 80$

III. EXPERIMENT ON THE STUDENT DATA

Based on Figure 1, experiment is conducted in 2 phase. First, the training and testing process using supervised Kohonen on the simulated data. Simulated data was generated using specification as stated in the previous section of this paper to simulate the real student data. Kohonen network need to be trained using large sample dataset to ensure its classification performance. Difficulties in getting the large student data especially in AHLS is the main constrain in this research. However, in training a neural network we must have a sample data to be trained with the network. So, we generate the student data and use it to train the supervised Kohonen network.

When the Kohonen network has been trained, the weights generated is stored and tested on a test data set. This is to see how much the network is able to classify the test data set. Eventually, the developed Kohonen network is improved by adjusting its learning parameters such as number of iteration, learning rate value, radius and etc to gain the best result.

In the second phase of our experiment, we use the real student data to be tested using Kohonen network. We let a class of students to use the learning system called SPATH which is developed by researchers under Software Engineering Department in Faculty of Computer Science and Information Systems, Universiti Teknologi Malaysia to collect the real student data. We manage to collect 31 data from the activity.

A. Training The Supervised Kohonen Network

At this stage, training process is conducted using three Kohonen networks which are: proposed supervised Kohonen network with hexagonal lattice, supervised Kohonen network with square lattice and unsupervised Kohonen network with conventional self-organizing learning algorithm.

We use 10-fold cross validation to validate our proposed model. The data set is divided into *k=10* sub sets, and the holdout method is repeated *10* times. Each time, one of the *10* subsets is used as the test set and the other 9 subsets are put together to form a training set. Then the average error across all *10* trials is computed. The advantage of this method is that it matters less how the data gets divided. Every data point gets to be in a test set exactly once, and gets to be in a training set *10-1* times. The variance of the resulting estimate is reduced as *k* is increased. A variant of this method is to randomly divide the data into a test and training set *k* different times. The advantage of doing this is that we can independently choose how large each test set is and how many trials we average over. In this experiment we use a total of 9000 data to be trained. Table II presents the proposed Kohonen network:.

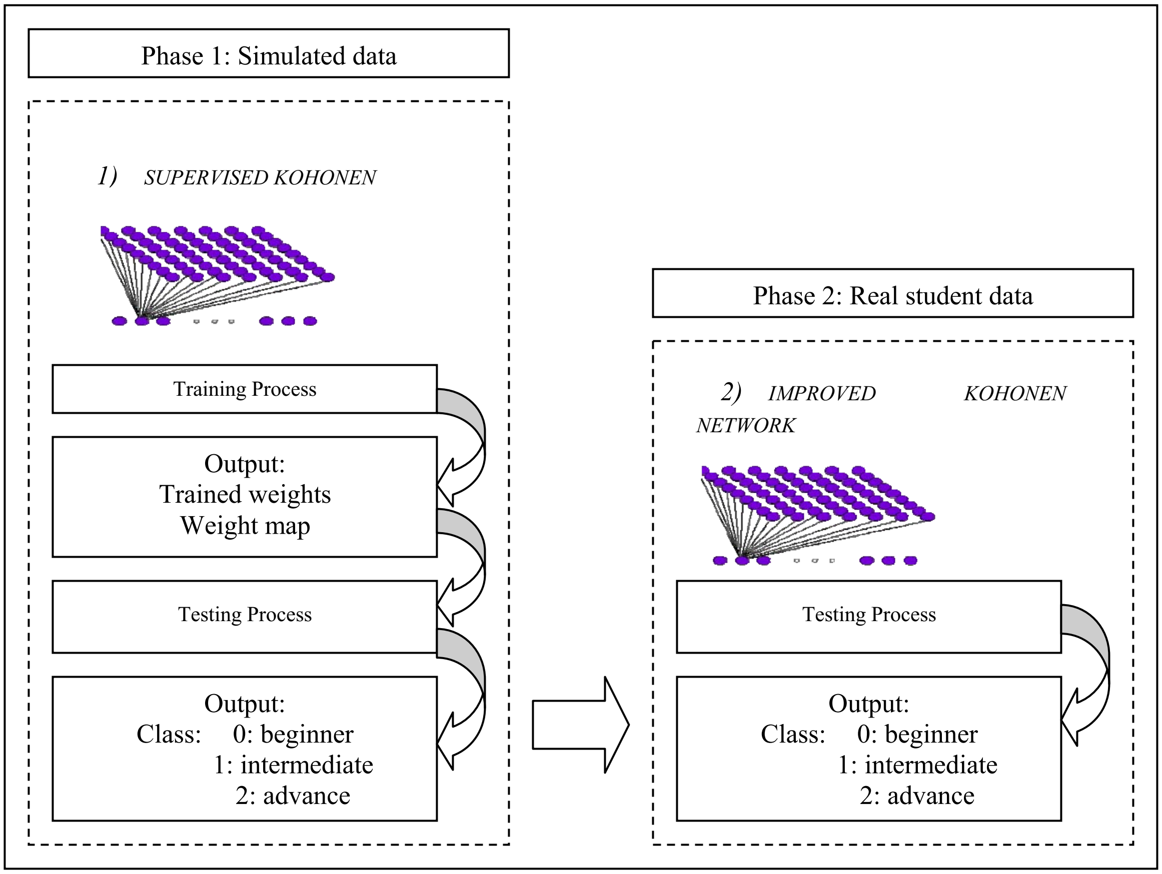

Figure 1. Supervised Kohonen Network Classifier Process

TABLE II. SUPERVISED KOHONEN NETWORK FOR TRAINING

Size	20 x 20
Dimension	2
Map structure	Rectangular grid
Lattice structure	Hexagon
Neighborhood function	Gaussian
Initial radius	10
Initial learning rate	0.5
Training sample size	8100 (Each sample has 9000-900 data)

During training, simulated data is divided into 10 samples to see the ability of proposed Kohonen network towards a different sample. Output generated during training is a set of weight values that has been updated to make it similar to the train data as much as possible. These weight values then organized in a map form so that it can be read and compared with the test data in the testing process afterward.

B. Testing The Supervised Kohonen Network

Weight values and map generated in the training process is compared with the test data to classify the data into 3 categories as discussed in the previous section. Test data generated from the simulation of the student data have no target data. It is a process to test the performance of proposed Kohonen network in classifying the data.

In this process, the network structure is the same but the test data sample is different as shown in Table III:

TABLE III. SUPERVISED KOHONEN NETWORK FOR TESTING

Size	20 x 20
Dimension	2
Map structure	Rectangular grid
Lattice structure	Hexagon
Neighborhood function	Gaussian
Initial radius	10
Initial learning rate	0.5
Number of iteration	1000, 3000 and 5000
Testing sample size	1-900,901-1800,1801-2700,2701-3600,3601-4500,4501-5400, 5401-6300,6301-7200,7201-8100,8100-9000

After that, results from each sample data is noted down and compared with the target data to get the accuracy percentage of the Kohonen network. At the end of the process, we test the trained weight values from each model to the real student data to see its performance in classifying real AHLS data.

C. *Result*

From the experiment, we find that supervised Kohonen network using hexagonal lattice structure excel in classifying the testing data with the highest accuracy compared with supervised Kohonen network using square lattice and conventional Kohonen network or SOM (Self-organizing map). Lattice structure determined how the neighborhood of the Kohonen map should be formed and the neurons in the map will be identified so that the weights associated with each neuron can be updated. Results from the experiment, the proposed Kohonen network classify all the test data samples with more than 80 % of accuracy and yield the overall accuracy with 81.3889 % (Table IV)

TABLE IV. ACCURACY TABLE

Data Sample	Percentage of Correct Classification		
	Hexagon	**Square**	**SOM**
1 - 900	83.6667	82.2222	71.0000
901 - 1800	80.4444	78.6667	71.5556
1801 - 2700	81.3333	80.0000	70.0000
2701 - 3600	82.0000	81.4444	71.0000
3601 - 4500	81.3333	80.0000	70.2222
4501 - 5400	80.6667	79.1111	70.2222
5401 - 6300	82.7778	82.1111	71.6667
6301 - 7200	79.7778	79.4444	67.8889
7201 - 8100	81.7778	80.2222	72.0000
8101 - 9000	80.1111	79.7778	72.1111
Average of Accuracy, %	81.3889	80.3000	70.7667

TABLE V. ERROR

Data Sample	Error		
	Hexagon	**Square**	**SOM**
1 - 900	0.1633	0.1778	0.2900
901 - 1800	0.1956	0.2133	0.2844
1801 - 2700	0.1867	0.2000	0.3000
2701 - 3600	0.1800	0.1856	0.2900
3601 - 4500	0.1867	0.2000	0.2978
4501 - 5400	0.1933	0.2089	0.2978
5401 - 6300	0.1722	0.1789	0.2833
6301 - 7200	0.2022	0.2056	0.3211
7201 - 8100	0.1822	0.1978	0.2800
8101 - 9000	0.1989	0.2022	0.2789
Average of error	0.1861	0.1970	0.2923

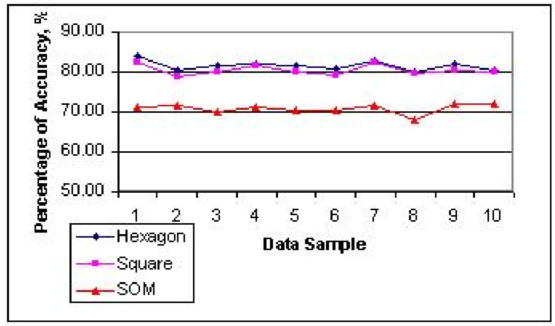

Figure 2. Graph of Accuracy

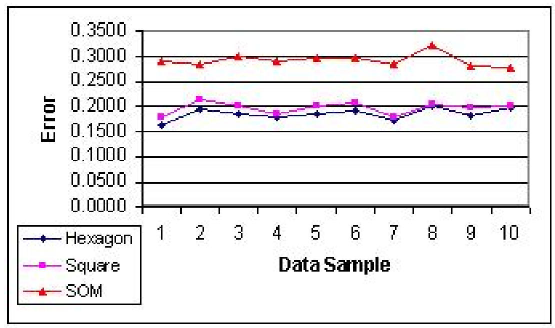

Figure 3. Graph of Error

Proposed Kohonen network also gives us the smallest average of error value, 0.1861 compared to the square supervised Kohonen (0.1970) and SOM (0.2923) as shown in Table V above.

Table VI shows the results yielded from the final phase of our experiment to the real student data. We take the trained weights from each Kohonen model to the real data and successfully prove that our proposed Kohonen network is the best classifier compared with the other 2 models.

TABLE VI. RESULTS FROM THE REAL STUDENT DATA

Kohonen Network	**Num of correct classification**	**Percentage of accuracy, %**	**Error**
Hexagon	16	51.6129	0.4839
Square	15	48.3871	0.5161
SOM	11	35.4839	0.6452

IV. CONCLUSION

The difficulties in identifying the student characteristics to determine their knowledge level is the key issue that always been a concern in AHLS researchers. We propose a supervised Kohonen network and improve its performance using hexagonal lattice structure to develop student model through classification of the student into 3 categories as stated before. Results from the experiment shows that the proposed network is capable of classifying the student data and gives us better results compared to the supervised Kohonen network using square lattice structure and conventional Kohonen network that use self-organizing learning algorithm. This study proposes an intelligence method to be applied in the AHLS environment to ease the process of understanding the student who is the main subject in the area. In this study, we use a simulated data to train the Kohonen network due to the constrains that we face in getting the real student data. In future, we hope to collect more real student data and try to use it to train the Kohonen network to improve and validate its performances toward real data.

V. REFERENCES

[1] Brusilovsky, P. (1999). Adaptive and Intelligent Technologies for Web-based Education. In: C. Rollinger and C. Peylo (eds.): Kunstliche Intelligenz, Special Issue on Intelligent Systems and Teleteaching. 1999 (4). 19-25.

[2] Zaiane, O. & Jun, L. (2001). Towards Evaluating Learner's Behaviour in a Web-based Distance Learning Environment. In: Proceedings of IEEE International Conference on Advanced Learning Technologies. 357-360.

[3] Esposito, F., Licchelli, O., & Semeraro, G. (2003). Discovering Student Models in E-learning Systems. In: Journal of Universal Computer Science. Vol. 10. No. 1 (2004). 37-47.

[4] Surjono, H.D. & Maltby, J.R. (2003). Adaptive Educational Hypermedia Based on Multiple Student Characteristics. In: W. Zhou et al. (Eds.): ICWL 2003, LNCS 2783. Springer-Verlag Berlin Heidelberg. 442-449.

[5] Siti Zaiton, M.H., Norazah, Y., Nor Bahiah, A., Paridah, S., & Zalmiyah, A.B., (2000). The Design and Implementation of an Adaptive Web-based Hypermedia Learning System: A Prototype. In: SPAtH Technical Report. 2000.

[6] Paridah, S., Nor Bahiah, H.A., Norazah, Y., Siti Zaiton, M.H., & Siti Mariyam, H.S. (2001). Neural Network Application on Knowledge Acquisition for Adaptive Hypermedia Learning. In: Chiang Mai Journal of Science. Vol 28, No. 1, June 2001.

[7] Catledge, L. & Pitkow, J.M. (1995). Characterizing Browsing Strategies in the World-Wide Web. In: Scientific Literature Digital Library, CiteSeer (IST).

[8] Mullier, D.J. (1999). The Application of Neural Network and Fuzzy Logic Techniques to Educational Hypermedia. School of Computing, Leeds Metropolitan University. PhD Thesis.

[9] Canter, D., Rivers R., & Storrs G. (1985). Characterising User Navigation Through Complex Data Structure. In: Behaviour and Information Technology. Vol. 4, Issue 2. 93-102.

[10] Zhou, Y., Freedman, R., Glass, M., Michael, J.A., Rorick, A.A. & Evens, M.W. (1999). Delivering Hints in a Dialogue-Based Intelligent Tutoring System. In: Proceedings of the 16th National Conference on Artificial Intelligent (AAAI_99). Orlando.