IJICTR International Journal of Information & Communication Technology Research

ITRC

Volume 4- Number 2- March 2012 (11-26)

E-Learners' Activity Categorization Based on Their Learning Styles Using ART Family Neural Network

Gholam Ali Montazer
Associate Professor, School of Engineering
TarbiatModares University
Tehran, Iran
montazer@modares.ac.ir

Hessam Khoshniat
MSc student, School of Engineering
TarbiatModares University
Tehran, Iran
hessamkhoshniat@gmail.com

Received: December 13, 2011- Accepted: February 22, 2012

Abstract—**Adaptive learning means providing the most appropriate learning materials and strategies considering students' characteristics. Grouping students based on their learning styles is one of the approaches which has been followed in this area. In this paper, we introduce a mechanism in which learners are divided into some categories according to their behavioral factors and interactions with the system in order to adopt the most appropriate recommendations. In the proposed approach, learners' grouping is done using ART neural network variants including Fuzzy ART, ART 2A, ART 2A-C and ART 2A-E. The clustering task is performed considering some features of learner's behavior chosen based on their learning style. Additionally, these networks identifythe number of students' categories according to the similarities among their actions during the learning processautomatically. Having employed mentioned methods in a web-based educational system and analyzed their clustering accuracy and performance, we achieved remarkable outcomes as presented in this paper.**

Keywords-component; Personalized e-Learning System; Adaptive Resonance Theory; ART Neural Network; Learning Style; Intelligent Tutoring System;

I. Introduction

Today's advances in information technology have shown great impacts on educational systems and adoption of e-learning systems to develop on-line courses is rapidly growing. To increase efficiency and effectiveness of the learning process, these systems should be to be able to provide personalized learning contents and to adapt strategy to support learner's diversity and individual needs. Personalization is often achieved by making a model of the learner's personality. This model can include personality factors, behavioral factors and knowledge factors [1].

Researchers have proposed various methods to automatically make a model of a learner. Some of these methods consider attributes like personality factors, behavioral factors, and so on. One of the approaches in student modeling is using data mining techniques such as clustering algorithms in order to discover patterns in user behaviors. It helps in categorizing them into some similar groups to choose the most relevant and specific learning materials and strategies for each student category. Most of these techniques consider features of one dimension of student factors e.g. they cluster students using only behavioral factors ignoring personality factors and so on. Learning style is one of the personality factors showing the way the student learns better. According to Felder and Silverman (1988), learners learn more effectively if they are provided by learning material which fits their learning styles. Therefore researchers emphasize on learner modeling based on learning style[2, 3].

Another shortcoming of these grouping methods is that they put learners into a predetermined number of clusters. In other words, although students and their characteristics in different courses may be very diverse, the clustering task groups them into a fixed number of clusters. This means that defining the right number of student clusters based on their various behavioral or personal factors is a problem.Furthermore, students may become more familiar with the web-based learning system and feel more or less comfortable with it during the learning process. As a result, their interactions, behavior, and therefore number of the clusters may change as the course goes on.

Dynamic identification of students' learning style unlike the static approach can adapt the system in different situation, while a static approach such as using questionnaires describes the learning style of a student at one specific point of time. Therefore, a dynamic approach can cope with extraordinary behavior of students and dynamically adapt the learning style of a student that changes over time.

In this paper, we address the aforementioned issues and propose a learners' grouping method. In the proposed methodology, students are clustered using their behavioral factors i.e. different aspects of their interactions when they work with the web-based education system. These behavioral features are chosen considering a learning style model. In other words, selected features are grouped based on different dimensions of the learning style proposed by Felder and Silverman. For the clustering task we employed four architectures of ART neural network including Fuzzy ART, ART 2A, ART 2A-C and ART 2A-E.We need a method in which the number of clusters is discovered wisely and automatically.These networks are able to learn and determine the number of clusters in this way.

We have evaluated functionality of the new methods on extracting students' groups from the underlying LMS legged data. The experimental results show that the match ratio between the obtained clusters using the proposed methodologies and preferences obtained by the questionnaires traditionally used for learning style assessment is high for the most of the dimensions of learning style. In addition, we have compared clustering precision of different types of ART-networks and the result shows that exploiting ART 2A-C neural network in grouping students results in more precision than Fuzzy ART, ART 2A-C, and ART 2A-E.

The rest of this paper is organized as follows. In Section 2, the literature review on various ways in learner/user modeling using clustering methods is presented. Background on learning styles is introduced in Section 3. Section 4 describes the proposed clustering approach: ART neural networks. In Section 5 modeling students' behavior using ART networks is given. Evaluation of the proposed method and the experimental results are discussed in Section 6 and conclusion is given in Section 7.

II. LITERATURE REVIEW

In the literature, different methods on learner modeling have been introduced. These techniques follow various approaches to make a model of learner in various aspects of behavior and knowledge such as his/her goals and plans, capabilities and previous knowledge, learning style and cognitive style, activities in the learning environment, attitudes or beliefs, and so on. The application which uses the model determines which parameters and dimensions of the learner's behavior will be considered to construct the learner's model. Since our work focuses on automatic grouping students based on their learning styles, a brief literature review of researches on these areas is presented.

In 1999, Carver and his co-workers developed an adaptive system which provided dynamic tailoring of the presentation of the course material based on the student's learning style detected using questionnaires [4]. In the same year Gilber and Han introduced Arthur, a Web-based instruction system that provides adaptive instruction based on student's learning style [5].

MAS_PLANG, a multi-agent system approach presented by Peña in 2002, employed an adaptation technique focusing on the suitable selection of didactic contents, navigation tools and navigation strategies depending on the student's learning style [6]. The critical issues influencing the adaptation mechanism based on the learning style information in Adaptive Educational Hypermedia Systems (AEHS) is considered in [7]. Cha proposed an intelligent learning system which detects learning styles by observing learner behavior patterns using decision tree approach [2]. In 2006, Graf and Kinshuk provided a basis for adaptivity by presenting a tool that enables learning management systems (LMS) to detect learning styles based on the behavior of learners during an online course [8]. Garcia and his co-workers in [3] evaluated the capability of Bayesian networks to model and detect students' learning style while they work with the Web-based education system. In [10], the learner model is extracted based on Felder–Silverman learning style model using the NBTree classification algorithm in conjunction with Binary Relevance classifier. The learners are classified based on their interests and their learning styles are detected using these classification results.

There exists a large number of clustering algorithm applications in e-learning environments in the literature and the choice depends on the particular application. An investigation of using Expectation Maximization (EM) to group learners into clusters using the data of their behaviors logged in the system was presented in [11].

The clustering task was performed to learn patterns reflecting his/her behaviors and construct groups of learners with similar behavior to provide an efficient collaborative environment. Mor and Minguillon proposed the use of clustering algorithms for grouping learners using information produced during learning process by the system such as user profile, navigational behavior, and academic results. They presented a framework to generate personalized itineraries for courses [12].

In [13] artificial neural network is employed to assess students and grade them into an ability level in a tutorial supervisor (TS) system. The system can grade both the student and the questions in a tutorial with minimal input from the human teacher using Kohonen's Self OgranizingMaps(SOM). Kohonen's SOM algorithm was also used in [14] to group similar learning materials according to their semantic relationships.

III. Learning Style

Intelligent educational environments put great importance on Learner’s characteristics and personality. They aimed at adjusting teaching strategy and customizing learning processes in such a way that they fit their learning styles as much as possible. In such systems, a model of student is built to provide personalized education which means the way each learner is trained and interacted will be different from other learners. This model provides valuable information about learners such as their previous knowledge, learning capabilities, interests, learning style, and learning process. The main role of the student model is to explain learners' behavior and personality so that the system can choose the most appropriate teaching strategy commensurate with their characteristics and requirements. Learning style is one of the learner's personal traits which has been subject of research from earlier times [15].

A learning style model places learners in some groups according to the ways they learn. The most well-known learning style models introduced into the literature are Myers–Briggs type indicator (MBTI) [16], Kolb’s model [17], Felder and Silverman learning style model (FSLSM) [9], Herrmann Brain Dominance Instrument (HBDI) [18], and Dunn and Dunn model [19].

Among these learning style models, we used FSLSM for engineering students. Because this model describes learning style deeply, some researchers believe that this model is the most suitable model which can be included in adaptive educational systems [4].

Felder-Silverman learning style model has four dimensions in the space of learning ways [20]. According to the perception dimension, which corresponds to the type of information the student preferably perceive, learners are categorized as sensory or intuitive learners. Sensory learners like experiments, sights, sounds, physical sensations, and obvious facts.

Table 1. Dimensions of Felder's learning style and their scales

Dimension	Scales
Perception	**Sensitive / Intuitive**
Input	**Visual / Verbal**
Processing	**Active / Reflective**
Understanding	**Sequential / Global**

Intuitive learners prefer theoretical information, abstracts, possibilities, insights, and so on.theInput dimension determines a kind of channel that information more effectively delivered to the learner through it. A visual learner learns more when they are given visual materials like pictures, diagrams, graphs, illustrations. Verbal learners remember better when they hear or read materials. They like listening, texts or sounds. The way the learners process information denotes the processing dimension. They may learn actively via experiments and with collaboration, engagement in physical activity or discussion. Reflective learners prefer to work alone, think about the information being presented to them and without trying things. The last dimension refers to the manner learners understand. Sequential learners solve problems step by step in a linear reasoning process. Global learners make a global picture of the concept, make intuitive leaps, and then work with details. Table 1 shows the dimensions of the Felder's learning style and their scales.

Richard Felder and Barbara Soloman in 1991 developed a questionnaire called Index of Learning Styles (ILSs) [21] to evaluate students' learning style on each dimension of the FSLSM [18]. Learners in adaptive educational systems are often asked to fill out this questionnaire at the beginning of online courses in order to determine students' leaning style. Then the outcome will be used to adjust teaching strategy and to provide learning contents and materials appropriately.

In our work, we used Felder's model to divide student's behavioral factors into some groups in such a way that the features related to a dimension are put into the same group. Then we applied our clustering algorithm to the students' logged data considering each group's features. In addition, to evaluate the accuracy of the proposed method, we used the ILSs questionnaire filled out by the learners participated in the experiment.

IV. Proposed Learners' Clustering Approach: ART Neural Networks

There exists a large number of clustering algorithms in the literature and the choice depends on the underlying learning system. For our purposes, we require an algorithm capable of assigning the students in an e-learning environment into some groups based on the similarities between patterns of their behavior and activities in the system. In other words, we need a method in which the number of clusters is discovered wisely and automatically. If the learners behave almost in the same way, the number of groups would be small. In contrast, groups of learners would be numerous if they follow different patterns of interactions with the system. In addition, since the clustering method is intended to be performed frequently as the course goes on, it should not be a time consuming algorithm.

ART neural network is a self-organizing clustering method that besides having some features like performance, economic usage of memory resources and temporal stability of stored knowledge [22], is also capable of recognizing the number of clusters automatically and simultaneously.

A. *Adaptive Resonance Theory (ART) Neural Network*

ART Neural Network has been developed by Carpenter and Grossberg in 1976 [23]. It was designed after the development of the *AdaptiveResonance Theory (ART).* It is said that ART networks solve the *stability-plasticity dilemma* which means the network is plastic enough to incorporate and learn new patterns while being stable enough to preserve previously learned information. In other words, they are always able to learn new patterns without forgetting the past [24].

The first version of ART, ART-1, proposed by Carpenter and Grossberg[25], is used to cluster binary data. Since then several variants of ART network have been developed. The most important ones are: ART-2 [26], an extension of ART-1, used to cluster analog data, ART-2A [27], an efficient version of ART-2 which is two to three orders of magnitude faster than ART-2, ART-2A-C[28], an extended version of ART-2A by complement coding, ART-2A-E [29], improved by employing Eucledian measurement of similarity, Fuzzy ART[29], a clustering method which accepts analog fuzzy input patterns ,ARTMAP [30], a supervised learning mechanism for binary data, and Fuzzy ARTMAP [31], which is a supervised learning algorithm for analog data. There are the other proposed variants of ART neural network too. Table 2 summarizes the history of ART neural network variants proposed into the literature.

B. *ART Neural Network Architectures and Learning Algorithms*

Figure 1 shows the common architecture of an ART network which consists of two layers of neurons. The input layer F_1 consists of M neuron and theoutput layer F_2 includes n neurons. If there exist a complement coding task as a preprocessing phase, Mwill the double of m, the input vector's dimension, otherwise it will be equal to m. The number of theoutput neurons, n, which refers to the number of clusters, is not predetermined and is set to 1 at the beginning of the training procedure and increases during the procedure. Each neuron in the F_1 layer is connected to every neuron on the F_2 layer through a weighted connection. Each neuron of F_2 represents a cluster recognized by the network. The weight vector, $\boldsymbol{w}_j$, corresponds to the j-th cluster where j, lying between 0 and n, is the index of the corresponding neuron in layer F_2.

B.1. *Learning Procedure*

Training procedure of an ART network consists of four main phases: *preprocessing*, *search* including *choice* and *match*, and *adaptation.*

a. Preprocessing

In this phase, some operations are performed on the input vectors, $A=(a_i)\ i=1,\dots,m$, to make them ready to feed into the network and be comparable with the prototypes stored in the synaptic network weights, wj. The elements of the vector should satisfy some constraints which depend on the type of the ART network. The resulting vector, $I=(i_k)\ k=1,\dots,M$, is fed to the F2 layer to be processed in the next choice phase.

b. Choice

The next step, is to estimate the similarity between the preprocessed input pattern and the synaptic weight vectors which are the clusters' prototypes learned by the network.As a summary, in this phase a choice function, T, is evaluated and the elements of the vector $\boldsymbol{T}=(T_1,\dots,T_n)$ can be seen as the result of comparison between the input pattern and the prototypes $\boldsymbol{w}_1=(w_{11},\dots,w_{1M}),\dots,$ $\boldsymbol{w}_n=(w_{n1},\dots,w_{nM})$. These prototypes are stored in the weights of the connections between F_1 and F_2 layers. The winner neuron with higher activity compared to the other neurons is then found:

$$J = \arg\max_j (T_j) \qquad j = 1,\dots, n \tag{1}$$

where$J$ is the index of the winning neuron.

c. Match

A vigilance criterion is evaluated once the winning neuron is selected. If the input pattern respects the vigilance test, it is added to the winning neuron's corresponding cluster. Otherwise, the output of the winning neuron is set to zero and the next neuron with the highest activity is selected to evaluate vigilance criterion. If none of the neurons pass the criterion, a new neuron is added to the F2 layer, which means a new cluster is created. Thus the vigilance parameter ρ controls the maximum size of clusters [32]. The greater vigilance parameter, the smaller size of clusters will be. If the similarity between the input pattern and the prototype represented by the best fitting output neuron is high enough, which means itshould be at least as high as the vigilance parameter, the pattern will be assigned to that cluster.

Table 2. Proposed ART neural network variants into the literature

ART Neural Network Version	Application
ART-1(G. Carpenter & Grossberg, 1987b)	Clustering
ART-2(G. Carpenter & Grossberg, 1987a)	Clustering
ART-2A-E(Pao, 1989)	Clustering
Fuzzy ART(G. Carpenter, et al., 1991)	Clustering
ART-2A(G. A. Carpenter, S. Grossberg, et al., 1991a)	Clustering
ARTMAP (G. A. Carpenter, S. Grossberg, & J. H. Reynolds, 1991)	Classification
Fuzzy ARTMAP (G. A Carpenter, et al., 1992)	Classification
Simplified Fuzzy ARTMAP (SFAM)(Kasuba, 1993)	Classification
Simplified ART (SART) (Baraldi & Parmiggiani, 1995)	Clustering
Adaptive Hamming Net(AHN) (Hung & Lin, 1995)	Clustering
ART-2A-C(Whiteley, et al., 1996)	Clustering
Gaussian ARTMAP(GA) (Williamson, 1996)	Classification
µARTMAP (Gomez-Sanchez, et al., 2002)	Classification

If the vigilance criterion is not complied, the winner neuron will be set to 0 and the next neuron with the highest activity will be chosen to be evaluated. If no similar enough cluster is found, a new cluster has to be created, where the current pattern is commonly used as the first prototype.

d. Adaptation

Once the vigilance criterion is satisfied which means a similar enough cluster is found, it will be adapted to the new cluster. In the most cases, the cluster prototype is slightly shifted toward the new input pattern. These two steps, *choice* and *match*, are repeated until the vigilance criterion is respected or all neurons on F_2 are evaluated and no cluster with enough similarity is found. If the similarity between the input pattern and the selected neuron is greater than the vigilance parameter, it is said that the Fuzzy ART network is in resonance. When it enters resonance, it learns the new input vector I by modifying the synaptic weight vector $\mathbf{w}_J$ of the neuron.

If all neurons on F_2 are evaluated and no prototype matches the input well enough, a new neuron is added to the F2 layer and the input is identically copied as the first prototype of a new cluster. Figure 2 demonstrates the mentioned training procedure.

In our system, we have employed four main types of ART neural networks which are used as clustering tools: Fuzzy ART, ART 2A, ART 2A-C, and ART 2A-E. These types of ART neural network are used to cluster analog patterns. ART 2A is an improved version of ART2, which work several orders of magnitude more efficiently than ART 2 [27]. So we did not use it in our system. Some features of the architectures described in the original publications are skipped, as they are not relevant for this analysis.

B.2. Fuzzy ART Neural Network

Fuzzy ART neural network, a modified version of the binary ART1, was introduced by Carpenter, Grossberg, and Rosenberg in 1991 [29]. Despite the ART1 which can only accept binary patterns, Fuzzy ART is able to accept analog fuzzy input patterns. Fuzzy ART is an unsupervised neural network for clustering patterns whose components are real numbers between 0 and 1.

The input layer F1 consists M where M is the double of m, the input vector's dimension. Initially, all the weight vectors' components are fixed to 1.

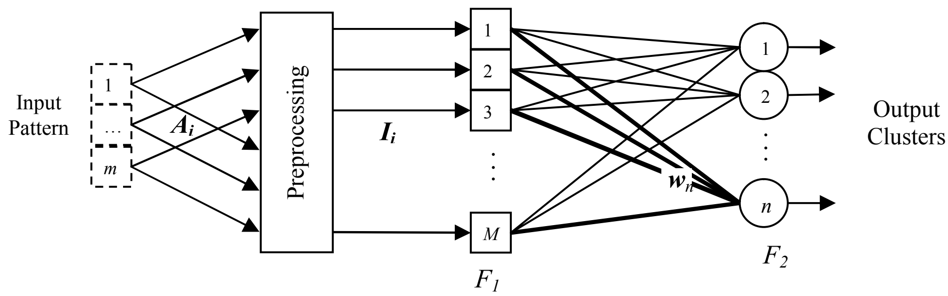

Figure1. Sample Fuzzy ART Neural Network

a. Preprocessing

Before feeding the input pattern to the network, it must be normalized into the range [0, 1]. Because Euclidean normalization causes loss of some information stored in the vector length of an input pattern, a modified normalization method called complement coding is used [22]. In this method, original vector can be coded into an input pattern by concatenating its complements to the original vector which doubles the dimension of input patterns:

$$I = (A, A^c) = (a_1, \ldots, a_m, 1-a_1, \ldots, 1-a_m)$$

$$a_i \in [0,1] \quad \forall i = 1, \ldots, m \tag{2}$$

The resulting vector is presented to layer F1. Therefore, M, the dimension of layer F1, is the double of m, the input vector's dimension.

b. Choice

The choice function, which should be evaluated for each neuron of F2, is defined by:

$$T_j = \frac{|I \wedge w_j|}{\alpha + |w_j|} \quad j = 1, \ldots, n \tag{3}$$

where | | is the L_1 norm of the vector, i.e. the sum of its components ($|p| \equiv \sum_{i=1}^{M} p_i$) , $\wedge$ is the fuzzy AND operator $(p \wedge q)_i \equiv \min(p_i, q_i)$ and $\alpha > 0$ is the choice parameter. The choice parameter, which provides a floating point overflow when $|w_j| \to 0$, is usually chosen close to zero for a good performance [31].

c. Match

The vigilance criterion is:

$$\frac{|I \wedge w_J|}{|I|} \geq \rho \tag{4}$$

(2)

where $\rho \in [0, 1]$ is the vigilance parameter and J is the index of the winning neuron on F2 determined by Eq. (1).

d. Adaptation

The adaptation rule is as follows:

$$w_J = \beta(I \wedge w_J) + (1-\beta) w_J \tag{5}$$

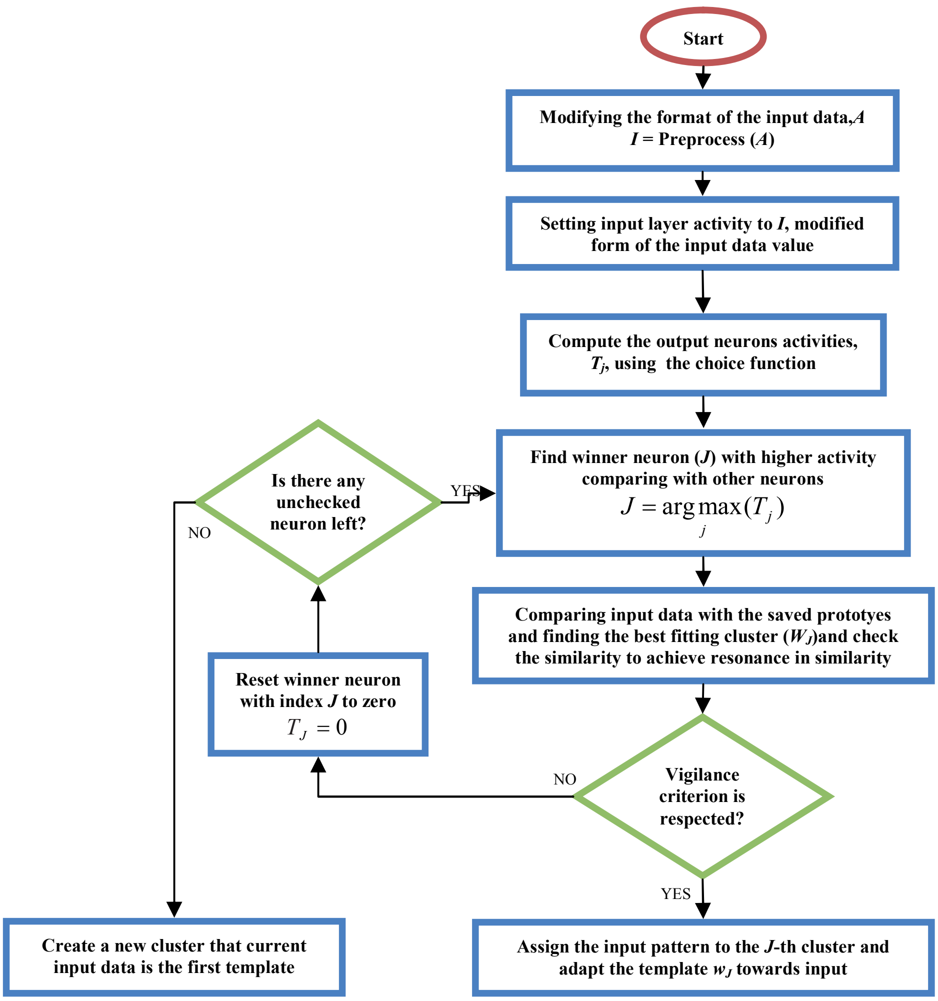

Figure 2. Learning Algorithm of ART Network

Here, $\beta \in [0,1]$ is the *learning rate* anddefines how quickly prototypes converge to the common minimum of all input patterns assigned to the same cluster.

B.3. ART 2A

ART 2 is proposed by Carpenter and Grossberg in 1987[26]. In contrast to the ART1 which only accepts binary patterns, ART2 network can learn analog input patterns. Four years later, they improved the order of magnitude of the learning process and introduced ART 2A[27]. ART 2A learning process is very similar to Fuzzy ART and the differences in the learning process are as follows.

a. Preprocessing

The elements of the input vectors should be greater than or equal to zero. They are normalized to unit Euclidean length:

$$I = \aleph(A) = \frac{A}{\|A\|}, \quad \|A\| = \sqrt{\sum_{i=1}^{m} a_i} \qquad a_i > 0, \|A\| > 0 \tag{6}$$

where$\aleph$ is the normalization function.

b. Choice

The choice function, which determines the bottom-up net activities, is defined by:

$$T_j = I.W_j, \quad j = 1,\ldots,n \tag{7}$$

The function defined here is slightly different from what is presented in [27] and is obtained by setting the choice parameter defined in original publication to zero(α=0).

c. Match

The vigilance criterion is defined by:

$$\rho \le I.W_J \tag{8}$$

where J is the index of winning neuron determined by Eq. (1).

d. Adaptation

Synaptic weights of the winning neuron are adapted through the below equation:

$$W_J^{(new)} = \aleph(\eta.I + (1-\eta).W_J^{(old)}) \quad 0 \le \eta \le 1 \tag{9}$$

WHERE η IS THE LEARNING RATE AND$\aleph$IS THE NORMALIZATION MENTIONED IN EQ. (6) .

B.4. ART 2A-C: Complement Encoding with ART 2A

Whitely added complement coding to the preprocessing phase of the learning algorithm to resolve the disadvantage of loss of all information coded in the length of an input pattern caused by normalization [28]. In other words, ART 2A cannot distinguish between two inputs A_1 and A_2, where A_1=$c.A_2$, with c>0[22].

a. Preprocessing

Complement coding is like what is done in Fuzzy ART:

$$I = (A, A^c) \quad a_i \in [0,1] \ \forall i = 1,\ldots,m \tag{10}$$

b. Choice

Despite ART-2, the normalization is embedded in the choice step:

$$T_j = \aleph(I).\aleph(W_j) \tag{11}$$

c. Match

The resonance occurs when the activation value of the winner neuron respects the vigilance test:

$$\rho \le T_J \tag{12}$$

where$J$ is the index of the winner neuron determined by Eq. (1).

d. Adaptation

Moving chosen prototype toward the accepted pattern is done by:

$$W_J^{(new)} = (\eta.I + (1-\eta).W_J^{(old)}) \quad 0 \le \eta \le 1 \tag{13}$$

whereη is the learning rate.

B.5. ART 2A-E

Another way to reduce the effect of normalization is introduced in [29]. Pao suggests replacement of the ART 2A-distance metric with a Euclidean measurement of similarity. The Euclidean algorithm used in this section is different from what introduced by Pao[22].

a. Preprocessing

All elements of the input pattern should lie in [0,1].

$$i_k \in [0,1] \quad \forall k = 1,\ldots,M \tag{14}$$

b. Choice

The choice function determines bottom-up net activities using Euclidean distance:

$$T_j = 1 - \sqrt{\frac{1}{M}\sum_{i=1}^{M}(i_i - w_{ji})^2} \tag{15}$$

c. Match

The vigilance test is as the same as Eq. (12).

d. Adaptation

Adaptation to the new pattern is also performed by Eq. (13).

The next section describes how we have employed, examined, and evaluated four types of ART neural networks as clustering methods to group students in an adaptive e-learning environment.

V. MODELING STUDENTS' BEHAVIOR USING ART NEURAL NETWORK

In order to cluster learners according to their behavioral factors, we have used logged data obtained from the underlying LMS, which includes students' interactions with the educational system. The features of the learners' behaviors we can record and measure generally depend on the functionality of the underlying Web-based education system. In our system we aim at grouping students based on their activities while working with the system such as using chat rooms, forums or type of learning materials they prefer like texts, videos, diagrams, and so on. The goal of this partitioning is to adopt the most appropriate teaching strategy according to the learners' requirements and characteristic. This strategy includes providing learning materials, adjusting sequence of learning objects, taking exams, generating recommendations, and so on.

Although we could simply apply an ordinary clustering algorithm such as k-means to the all data extracted from log files by taking all features of the student behavior into account and assigning them to some groups, there exists some shortcomings on this method. In such clustering algorithms the number of clusters should be specified before performing the task. Since in applications like our learners grouping task which almost everything like courses, students and their behavior is changing dynamically and therefore groups of students are always updating, use of these classic clustering methods may not be efficient. A possible solution is to add a preprocessing phase before performing the clustering task to find the right number of groups. Since the grouping of student is frequently done as the course goes on and the clustering task should not be time consuming, employing another method to determine the right number of clusters may not be reasonable. It will be better that the number of learners' groups is recognized by the clustering method automatically. In addition, if we consider all features together, some of dependencies, relationships and even contrasts and conflicts between the features may not result in the best clusters.

The next subsections describe how the mentioned issues have been resolved in our work.

A. Grouping Features

In order to consider relationships between behavioral factors and student's learning style and eliminate the aforementioned issue, we divide these features into four groups according to their relationship with dimensions of the Felder's learning style model. So instead of grouping students based on all behavioral features, we cluster them into some groups in each dimension and adapt the learning strategies and material according to the clusters the learner is assigned to. That is to say a student will be assigned to four groups, corresponding to the Felder's learning style model. This helps us to dynamically define our learning policies based on the groups detected in the dimension. Table 3 lists the behavioral factors considered for the clustering task in each dimension.

B. Determining the number of student clusters

Identification of the number of students' clusters is automatically performed by the ART network. Adjusting the vigilance parameter in the network, would make it possible to find the right number of clusters in each dimension. This capability of the network in identifying the number of clusters based on the similarity level between learners' removes the dependency of the system to students' personal factors. In other words, the system does not have to change its parameters (i.e. number of groups in each dimension) whenever the students change their behavior or the course starts with new students.

VI. EVALUATION OF THE PROPOSED METHOD

To evaluate the performance and accuracy of the proposed method, we conducted a study on 98 students who participated in an online undergraduate course about fundamentals of computer networks. This course seemed to be appropriate for our purpose because we needed a course which had both theoretical and practical subjects, diagrams and visual materials besides textual contents.

The online course learning contents contained all kinds of material including texts, audio lectures, videos, and slideshows. Each course chapter had an abstract section, an outline, an introduction, a chapter's content section, assignments which all students had to do, optional exercises, an end of the chapter exam, and a class project. Chapter contentswere available in text, voice lectures, videopresentations, and slideshows.

Table 3.Behavioral factors considered for the clustering task in each dimension

Dimension	Behavior Observed
Perception	**Type of reading material (abstract, concrete)** **Type of slideshows** **Dedicating time for reading concepts and theories** **Dedicating time for reading examples and facts** **Doing additional exercises** **Reading additional examples** **Exam doing and revision time**
Processing	**Participation in forums** **Use of chat and mail systems** **Participation in collaborative/group tasks** **Choosing group or individual class projects**
Input	**Listening to lectures** **Using video materials** **Reading textual materials** **Type of slideshows** **Using diagrams and charts**
Understanding	**Pattern of access to the course materials** **Dedicating time for lessons' objectives and overviews**

The assignments and projects are supposed to be done in groups of two or individually. At the end of the course, each student will have to pass a written exam. In addition, mail and chat system, forums, and discussion rooms were provided to enhance the students' collaborations. The course was designed so that all students needed to learn everything and they were examined on all topics. Therefore the course was suitable for investigation of individual learning.

The whole course was managed via a customized LMS which logged all of the students' interactions with it. In addition, in order to make sure the approach is correct, we asked the students to fill out the ILS questionnaire to get information about their learning styles.

Information about employing different types of ART neural networks in clustering students, evaluation method, and the result of the evaluation is presented in the following subsection.

A. System architecture

The architecture of the web-based educational system we have used is demonstrated in Figure 3. As shown in this figure, the system consists of four main components: user interface, student model, teaching model, and domainknowledge. Web-based user interface provides students with learning materials, assignments, exams and recommendations. Student's model contains four subsystem including students' profile database, students' interactions logger, students' log files analyzer, and clustering module. Recommender subsystem and adaptive content provider constitute the teaching module. Finally learning materials are provided by the domain knowledge.

Figure 4 illustrates structure of the clustering module which consists of four ART networks; each of them corresponds to one of the Felder's learning style dimensions. Log file analyzer unit does a preprocessing task on the students' interactions to extract information usable for the clustering unit. Grouping of the students is performed frequently by the clustering unit as the course goes on and their clusters are dynamically updated. In the next subsection the way the proposed method is evaluated is described.

B. Method of evaluation

To evaluate the proposed method, the results obtained from automatic students' grouping are analyzed considering the results obtained from the ILS questionnaire filled out by the students at the start of the online course. By calculation of the ILS preference, each student is given a grade form the 0 to 11 grade range on each dimension of Ferler's learning style. We have compared result of the ILS filled out by the students who were assigned to the same cluster in each dimension. Moreover, we have calculated average and standard deviation of the ILS result of the students belonging to the same cluster to find out how similar they are in terms of learning style and behavior. We have carried out the experiment using aforementioned types of ART networks including Fuzzy ART, ART 2A, ART 2A-C, and ART 2A-E in the clustering module and a deep comparison over their results, performance, and their accuracy in grouping students is done. Structure of the tentative clustering module contained four networks working in parallel is demonstrated in Figure 5. As shown in this figure, the module consists of four units. Each unit, which corresponds to one of Felder's learning style dimensions, uses the four types of ART network to recognize the student's group in that dimension.

The analysis results are discussed in the following subsection.

C. Results

We employed four types of ART neural networks in the clustering module of our system. Logged data of the students' activities in the system were preprocessed to comply with the constraints of the ART networks. Preprocessed students' data was fed to the four units; each of them corresponded to one of Felder's learning style model. In each unit, we clustered students employing Fuzzy ART, ART 2A, ART 2A-C, and ART 2A-E in order to compare their clustering accuracy. Analysis of the result is done by calculating average and standard deviation in the value of the ILS result of the students who belong to the same cluster. In other words, we have considered the similarities of the students assigned to the same cluster using the value obtained by the ILS questionnaire.

The result of clustering students considering features related to the perception dimension using Fuzzy ART, ART 2A, ART 2A-C, and ART2A-C networks is presented in Table 4, Table 5, Table 6, and Table 7, respectively. The ILS result of the students and the number of students assigned to each cluster are shown in these tables. μ_k andσ_k, which are used in the tables, are average and standard deviation of the ILS result of students assigned to cluster k, respectively:

$$\mu_k = \frac{1}{n}\sum_{i=1}^{n_k} ILS_i \quad i = 1,\dots,n_k\,, ILS_i = 1,\dots,11 \tag{16}$$

$$\sigma_k = \sqrt{\frac{1}{n}\sum_{i=1}^{n_k}(ILS_i - \mu)^2} \tag{17}$$

wheren_k is the number of students assigned to the cluster k and ILS_i is the result of ILS questionnaire filled out by the student i.

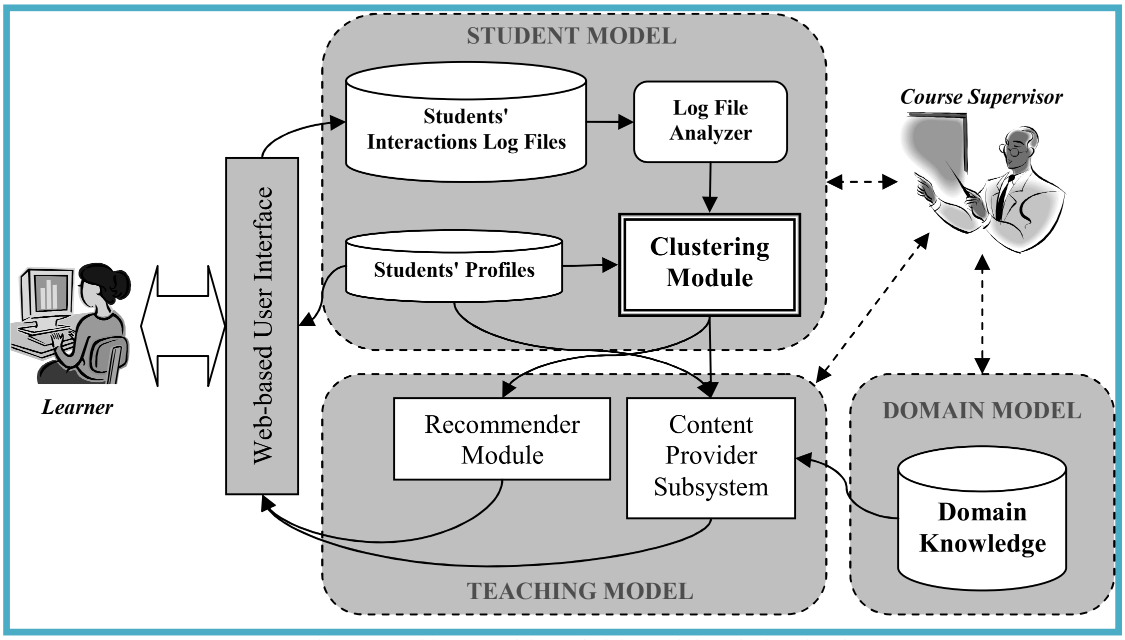

Figure 3. Architecture of the underlying web-based educational system

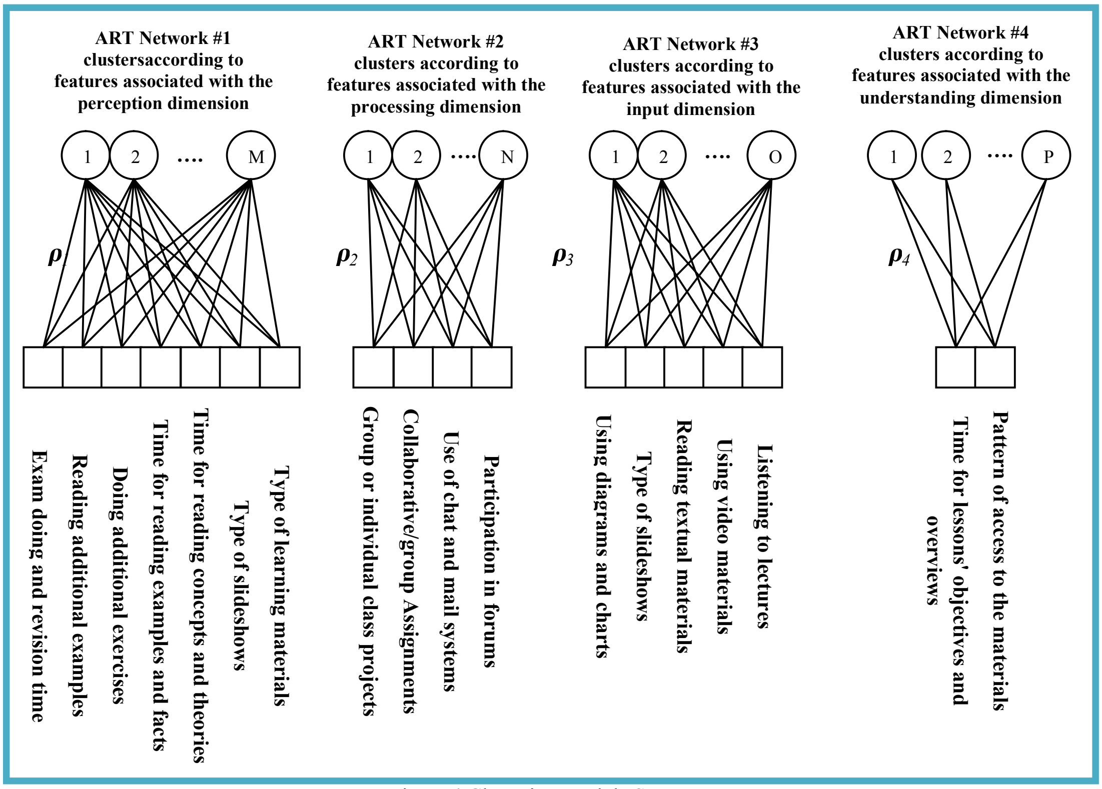

Figure 4.Clustering Module Structure

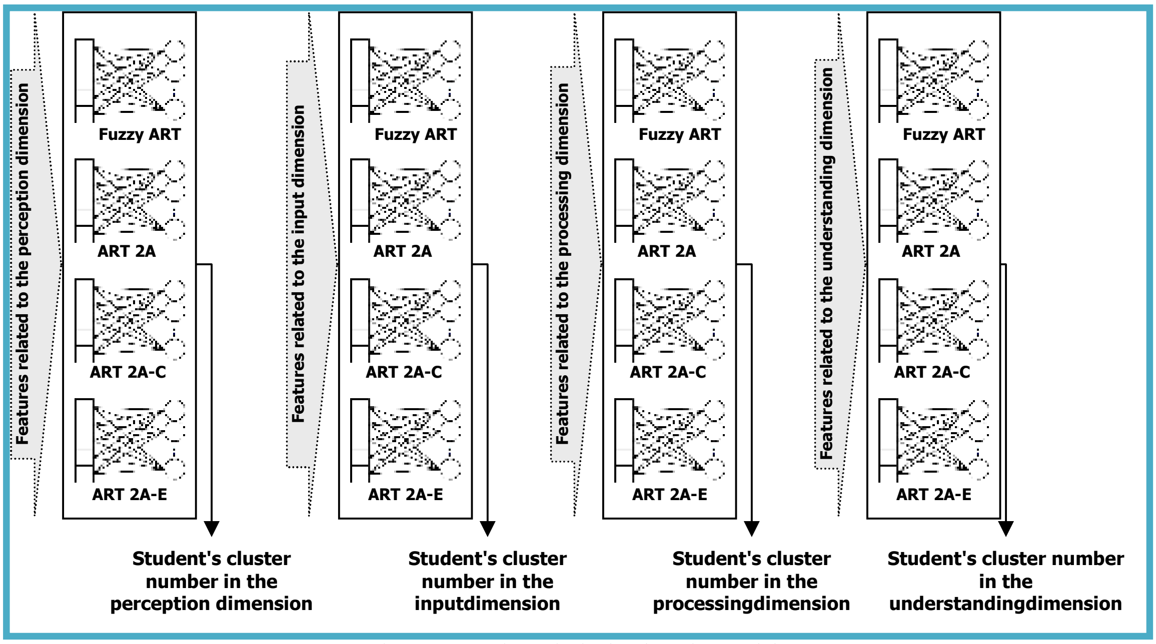

Figure 5.Tentative units in the clustering module

We also use two more symbols, M_{dim} and S_{dim}, which are calculated using the following equations:

$$M_d = \frac{1}{K}\sum_{k=1}^{K}\sigma_k \quad d \in \{perception, input, processing, understading\} \tag{18}$$

$$S_d = \sqrt{\frac{1}{K}\sum_{k=1}^{K}(\sigma_k - M_d)^2} \tag{19}$$

where$K$ is the number of clusters created in the dimension d. In fact, M_d represents the average of the standard deviations in the ILS result of the students for all clusters and S_d shows the deviation in σ_k of all clusters. In other words, M_d shows how much the clusters contain students with similar learning style and S_d describes how much the clusters vary in the quality defined by M_d.

Before applying the clustering methods, we analyzed the result of ILS questionnaires filled out by the students participating in the course. The analysis showed that distribution of the students in all dimensions is almost uniform. Therefore a method will be appropriate for our purpose that creates clusters with two conditions: their prototypes cover the range [0, 11] and also variance of the clusters' members (σ_k) is small. Since the students' distribution is almost uniform, creating two or more cluster with close μ_k indicates the network is not able to recognize groups of students properly.

As can be seen in Table 4, most of the students assigned to the same cluster which is created using Fuzzy ART network, have almost close values on the ILS result in the perception dimension.As shown in Table 5 ART 2Ain the perception dimension formed three clusters with only one member and two large clusters with 31 and 59 students assigned to. Cluster 1 contains students with the ILS result from 2 to 10 which means the network put students with different learning style in the same cluster. The standard deviation of *ILS* value of the clusters' members, $\boldsymbol{\sigma_k}$, has a great value compared to the clusters created by Fuzzy ART network. Therefore, ART 2A resulted in great values for $M_{perception}$ and $S_{perception}$ in comparison with the ones produced by Fuzzy ART.

Table 6 presents the result of employing ART 2A-C network in clustering students in the perception dimension. As shown in this table, clusters produced by this network have small variances in the *ILS* values of the students assigned to. It means complement coding phase has a positive effect on the performance ART 2A-C. Moreover, considering information presented in Table 4, Table 5, and Table 6, it can be concluded that ART 2A-C can group students better than Fuzzy ART and ART 2A networks. It resulted smaller values in both $M_{perception}$ and $S_{perception}$ than the two other networks.

Employing ART 2A-E neural network resulted in a similar outcome to ART 2A-C network. As shown in Table 7, this network achieved 0.88 and 0.27 for $M_{perception}$ and $S_{perception}$, respectively. Although ART 2A-E resulted in smaller value for the average of $\boldsymbol{\sigma_k}$of the clusters than ART 2A-C, the produced standard deviation of $\boldsymbol{\sigma_k}$is greater.

Table 8 summarizes the result of using the four networks in grouping students in the perception dimension. As can be seen, ART 2A-C and ART 2A-E networks produced smaller values for $M_{perception}$ and $S_{perception}$ than Fuzzy ART and ART 2A. The first two networks resulted in more consistent and homogenous clusters. Fuzzy ART created 2 clusters with μ_k almost equal to 9.00 which means we have two different clusters contain similar students in learning style. ART 2A behaved worse as it created three clusters with μ_k equal to 2.00. In contrast, ART 2A-C and ART 2A-E produced clusters so that their prototypes cover the range [0, 11]. Since the recommendations and learning contents in the system are generated according to the prototypes of the clusters, consistent clusters with small value in σ_k result in more well-matched personalization efforts with the students' needs and characteristics.

Table 4.Result of clustering the students' activities data related to the perception dimension using Fuzzy ART

Cluster Number	ILS result of the students assigned to the cluster (ILS_i i=1,…,n)	n_k	σ_k	μ_k
Cluster 1	10, 7, 5, 10, 8, 7, 9, 7, 7, 8, 8, 7, 8, 7, 7	15	1.29	7.67
Cluster 2	3, 5, 5, 4, 5, 4, 2, 3, 5, 3, 3, 4, 4, 3, 4, 4, 5, 3, 3, 3	20	0.91	3.75
Cluster 3	0, 2, 1, 0, 2, 2, 2, 0, 2, 0, 2, 2, 2, 3, 2, 2, 0, 2, 3	19	1.02	1.53
Cluster 4	9, 10, 8, 10, 10, 10, 9, 8, 10, 9, 9, 9, 8, 9, 9, 10, 9, 9, 9, 9, 11, 10, 10	23	0.76	9.30
Cluster 5	7, 5, 5, 6, 4, 6, 5, 6, 7, 5, 7, 6, 6, 6, 6, 5	16	0.86	5.75
Cluster 6	8, 8, 11, 8, 11	5	1.41	9.00
	$M_{perception}$		1.04	
	$S_{perception}$		0.26	

Table 5.Result of clustering the students' activities data related to the perception dimension using ART-2A

Cluster Number	ILS result of the students assigned to the cluster(ILS_i i=1,…,n)	n_k	σ_k	μ_k
Cluster 1	10, 3, 7, 5, 10, 8, 5, 9, 5, 7, 9, 10, 8, 2, 4, 10, 10, 5, 10, 1, 4, 9, 2, 2, 0, 8, 5, 10, 5, 5, 9, 6, 7, 9, 7, 4, 9, 6, 8, 5, 3, 2, 8, 3, 6, 2, 3, 4, 4, 8, 6, 8, 6, 10, 2, 3, 10, 8, 2	59	2.88	6.03
Cluster 2	0, 0, 7, 3, 0	5	3.08	2.00
Cluster 3	2	1	0.00	2.00
Cluster 4	2, 2, 8, 4, 9, 9, 7, 5, 10, 9, 7, 9, 7, 4, 6, 8, 3, 9, 6, 5, 7, 9, 11, 11, 3, 10, 2, 3, 7, 3, 5	31	2.81	6.45
Cluster 5	2	1	0.00	2.00
Cluster 6	0	1	0.00	0.00
	$M_{perception}$		1.46	
	$S_{perception}$		1.60	

Table 6.Result of clustering the students' activities data related to the perception dimension using ART-2AC

Cluster Number	ILS result of the students assigned to the cluster(ILS_i i=1,…,n)	n_k	σ_k	μ_k
Cluster 1	10, 10, 8, 9, 9, 10, 8, 10, 10, 10, 9, 8, 10, 9, 7, 9, 7, 9, 8, 8, 8, 9, 9, 10, 9, 9, 8, 9, 8, 9, 11, 10, 11, 10, 10	35	1.01	9.09
Cluster 2	3, 0, 2, 1, 0, 2, 2, 2, 3, 2, 0, 2, 0, 2, 2, 3, 2, 3, 2, 3, 3, 3	22	1.06	1.91
Cluster 3	7, 7, 7, 5, 5, 5, 6, 4, 6, 5, 6, 7, 5, 7, 7, 6, 6, 6, 5	19	0.94	5.89
Cluster 4	5, 5, 5, 4, 5, 4, 3, 4, 4, 3, 4, 4, 5, 3	14	0.77	4.14
Cluster 5	6, 8, 7, 8, 7	5	0.84	7.20
Cluster 6	2, 0, 2	3	1.15	1.33
	$M_{perception}$		0.96	
	$S_{perception}$		0.14	

Table 7.Result of clustering the students' activities data related to the perception dimension using ART-2AE

Cluster Number	ILS result of the students assigned to the cluster(ILS_i i=1,…,n)	n_k	σ_k	μ_k
Cluster 1	10, 10, 8, 9, 9, 10, 8, 10, 10, 10, 9, 8, 10, 9, 7, 9, 7, 9, 8, 8, 8, 9, 9, 10, 9, 9, 8, 9, 6, 8, 8, 7, 8, 7	34	1.07	8.62
Cluster 2	3, 2, 2, 3, 2, 2, 2, 2, 2, 2, 2,	11	0.40	2.18
Cluster 3	7, 5, 5, 5, 4, 5, 4, 3, 3, 4, 3, 4, 3, 4, 5, 3, 3, 3, 3,	19	1.11	4.00
Cluster 4	0, 2, 1, 0, 2, 0, 0, 2, 0,	9	0.97	0.78
Cluster 5	7, 7, 5, 5, 5, 6, 4, 6, 5, 6, 4, 7, 5, 7, 7, 6, 6, 6, 5,	19	0.99	5.74
Cluster 6	9, 11, 10, 11, 10, 11	6	0.75	10.17
	$M_{perception}$		0.88	
	$S_{perception}$		0.27	

Table 8.Summarized result of students' clustering in the perception dimension by the four networks

Cluster Number	Fuzzy ART			ART 2A			ART 2A-C			ART 2A-E		
	n_k	μ_k	σ_k	n_k	μ_k	σ_k	n_k	μ_k	σ_k	n_k	μ_k	σ_k
Cluster 1	15	7.67	1.29	59	6.03	2.88	35	9.09	1.01	34	8.62	1.07
Cluster 2	20	3.75	0.91	5	2.00	3.08	22	1.91	1.06	11	2.18	0.40
Cluster 3	19	1.53	1.02	1	2.00	0.00	19	5.89	0.94	19	4.00	1.11
Cluster 4	23	9.30	0.76	31	6.45	2.81	14	4.14	0.77	9	0.78	0.97
Cluster 5	16	5.75	0.86	1	2.00	0.00	5	7.20	0.84	19	5.74	0.99
Cluster 6	5	9.00	1.41	1	0.00	0.00	3	1.33	1.15	6	10.17	0.75
$M_{perception}$			1.04			1.46			0.96			0.88
$S_{perception}$			0.26			1.60			0.14			0.27

As can be seen in table 8, the first cluster took students with high value for *ILS* which means they are sensitive learners. But these clusters have almost great value for σ_k showing some intuitive learners are also assigned to these clusters. Having analyzed logged data of these students, we discovered that some intuitive learners really behaved like sensitive learners. This is due to the course nature which needs "doing" more than "thinking". Although some students are intuitive learners, they preferred doing practical exercises and the system put them beside other sensitive learners and adjusted the learning contents and recommendations based on their clusters' prototypes which are closer to their characteristics.

Result of applying the four networks to the students' logged data associated with the processing dimension is presented in Table 9. As shown in this table, ART 2A-C and ART 2A-E achieved smaller values for$M_{perception}$ and $S_{perception}$compared to the other networks. Like the perception dimension, ART 2A created two clusters with only 1 and 2 members and another one with 50 members (more than half of the students are assigned to this cluster). It indicates that ART 2A is not appropriate for employing in the clustering module**.**

Grouping students using their logged data related to the input dimension of Felder's learning style model is also performed using Fuzzy ART, ART 2A, ART 2A-C, and ART 2A-E and the result is presented in Table 10. As can be seen, again ART 2A-C network brings about the best result in producing clusters among the other networks. Even though in this dimension all networks attained almost similar result for $M_{perception}$ and $S_{perception}$, there are some points that make some differences between networks' performance. ART 2A-E created two clusters with $\mu_k$equal to 2.00 and 2.25 which mean it recognizes two clusters with the same characteristics for their members. In addition, Fuzzy ART formed two clusters containing 10 and 8 members all of whom have $\mu_k$is equal to 7.20 and 7.75, respectively. Again, we have two distinct clusters with similar students. Since students' distribution is uniform, creating two or more cluster with close μ_k indicates the network is not able to recognize groups of students properly.

Table 11 illustrates performance of the four networks in clustering students in the understanding dimension. Although Fuzzy ART reached the smallest value for σ_k among the networks, it is not suitable for clustering. The small value is due creation of one cluster with only one member and as a result variance equal to 0. Fuzzy ART network created two clusters with 47 and 50 members and another cluster with just one student assigned to. This shows Fuzzy ART cannot group the students properly. Among the other three networks, ART 2A-C resulted in the smallest value for $M_{understanding}$ and $S_{understanding}$. For this dimension, the approach yields less accurate results compared to the other dimensions which can be explained by the overlapping of patterns within groups caused by the small number of features we considered for this dimension. Therefore, further investigations are necessary, dealing with the extension of this method on this dimension.

A summarization of the results presented in Table 7, Table 8, Table 9, and Table 10 and discussed before, is shown in Table 12 and Table 13. In Table 12, $M_{\{dim\}}$of the clusters formed by the four networks in all dimensions is illustrated. As can be concluded from the information presented in this table considering all dimensions, ART 2A-C is able to construct groups of students in which the members have high similarity in their learning styles. Moreover, the clusters produced by ART 2A-C are more consistent as they have smaller value for $S_{\{dim\}}$for all dimensions as shown in Table 13. In other words, there exists small variance in σ_k of all clusters created by ART 2A-C**.** A comparison of capability of the networks in grouping students is demonstrated in Figure 6. In this figure, average of M_{dim}andS_{dim}presented in Table 12 and Table 13 is compared. As can be seen, considering all dimensions, ART 2A-C showed more accuracy than the other networks.

Table 9.Result of student clustering in the processing dimension by the four networks

Cluster Number	Fuzzy ART			ART 2A			ART 2A-C			ART 2A-E		
	n_k	μ_k	σ_k	n_k	μ_k	σ_k	n_k	μ_k	σ_k	n_k	μ_k	σ_k
Cluster 1	36	8.86	1.20	36	5.67	3.34	27	8.48	1.45	35	7.77	1.91
Cluster 2	21	3.05	1.36	2	2.00	1.41	21	1.62	0.97	14	1.43	1.09
Cluster 3	13	1.92	1.66	50	6.10	2.63	23	4.39	0.66	15	3.20	1.32
Cluster 4	24	5.29	1.04	9	6.00	3.08	15	6.07	1.22	20	5.00	1.12
Cluster 5	4	8.00	2.00	1	0.00	0.00	12	9.33	1.07	14	9.07	1.21
$M_{processing}$			1.45			2.09			1.08			1.33
$S_{processing}$			0.38			1.38			0.30			0.34

Table 10.Summarized result of students' clustering in the input dimension by the four networks

Cluster Number	Fuzzy ART			ART 2A			ART 2A-C			ART 2A-E		
	n_k	μ_k	σ_k	n_k	μ_k	σ_k	n_k	μ_k	σ_k	n_k	μ_k	σ_k
Cluster 1	32	8.78	1.48	16	5.50	1.21	18	5.39	1.20	16	5.69	1.20
Cluster 2	18	2.89	1.41	22	7.91	1.31	16	9.81	1.11	27	7.07	1.11
Cluster 3	10	7.20	1.14	15	10.07	0.80	26	2.00	1.13	20	9.95	0.83
Cluster 4	20	4.70	1.22	27	2.11	1.25	19	8.11	1.37	15	3.27	1.58
Cluster 5	10	1.30	0.67	11	6.55	1.13	11	7.00	0.63	16	2.25	1.48
Cluster 6	8	7.75	1.75	7	4.57	0.79	8	4.63	0.74	4	2.00	0.82
M_{input}			1.28			1.08			1.03			1.17
S_{input}			0.37			0.23			0.28			0.32

Table 11.Summarized result of students' clustering in the understanding dimension by the four networks

Cluster Number	Fuzzy ART			ART 2A			ART 2A-C			ART 2A-E		
	n_k	μ_k	σ_k	n_k	μ_k	σ_k	n_k	μ_k	σ_k	n_k	μ_k	σ_k
Cluster 1	47	2.45	1.74	65	5.85	3.07	34	4.62	1.81	34	2.09	1.85
Cluster 2	50	7.56	1.81	30	4.00	2.70	35	8.37	1.50	27	5.15	2.16
Cluster 3	1	9.00	0.00	3	0.67	0.58	29	1.79	1.52	37	7.89	1.81
$M_{understanding}$			1.18			2.12			1.61			1.94
$S_{understanding}$			1.03			1.35			0.17			0.19

Table 12. Comparison of $M_{\{dim\}}$ of the clusters in all dimensions

	Fuzzy ART	ART 2A	ART 2A-C	ART 2A-E
$M_{Perception}$	1.04	1.46	0.96	0.88
M_{Input}	1.28	1.08	1.03	1.17
$M_{Processing}$	1.45	2.09	1.08	1.33
$M_{Understanding}$	1.18	2.12	1.61	1.94
Average	*1.24*	*1.69*	*1.17*	*1.33*

Table 13. Comparison of $S_{\{dim\}}$ of the clusters in all dimensions

	Fuzzy ART	ART 2A	ART 2A-C	ART 2A-E
$S_{Perception}$	0.26	1.60	0.14	0.27
S_{Input}	0.37	0.23	0.28	0.32
$S_{Processing}$	0.38	1.38	0.30	0.34
$S_{Understanding}$	1.03	1.35	0.17	0.19
Average	*0.51*	*1.14*	*0.22*	*0.28*

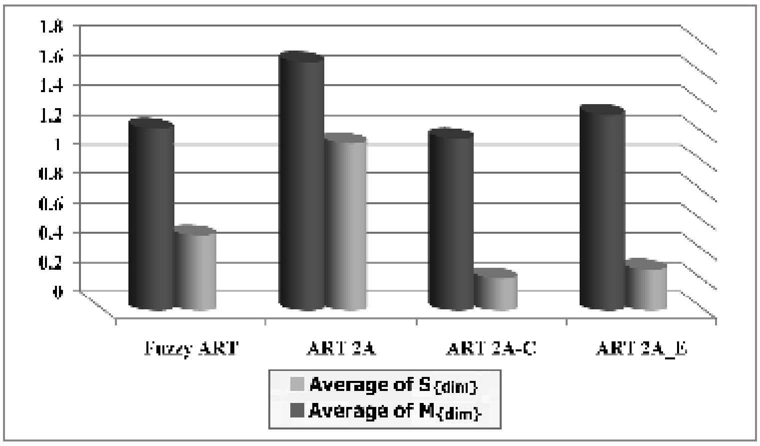

Figure 6.Comparison of ART neural network variants in students clustering

VII. CONCLUSION

In this paper, an application of ART neural networks in an online educational system to make a model of students was presented. In this system, we group students based on some of their behavioral factors and interactions with the system in order to adopt the best and the most appropriate learning materials. In addition, since there are some relationships and dependencies between these behavioral features, we divided the features into four groups according to their associations with the dimensions of Felder's learning style model. In this approach, the right number of students' clusters is identified automatically by the ART network.

To evaluate performance and accuracy of the proposed method, we conducted a study on 98 students participated in an online undergraduate course. In the proposed approach, this learners' grouping is done employing different ART neural network architectures including Fuzzy ART, ART 2A, ART 2A-C, and ART 2A-E. These methods automatically identify the number of students' categories according to the similarities among their actions during the learning process. At the start of the course, students were asked to fill out the ILS questionnaire in order to evaluate the clusters they are assigned to. We have analyzed the clusters calculating average and standard deviation of the ILS result of the students assigned to the clusters. Having compared clustering accuracy of the networks together, we observed that ART 2A-C neural network has more precision and more accuracy in putting students into some groups according to their behavior in the system. It formed clusters in such a way that the members have high similarity in their learning style. Since the recommendations and the learning strategies in the system are adjusted according to the clusters' prototypes, assigning students to clusters properly helps system provide more personalized content and learning recommendations.

VIII. ACKNOWLEDGEMENT

The authors would like to thank Mr.Vahid Fathi for his precious assistance in progression of the research.

REFERENCES

[1] Gu, Q., & Sumner, T. (2006). Support personalization in distributed e-learning systems through learner modeling. *Information and Communication Technologies, 2006.ICTTA'06. 2nd, 1*, 610-615.

[2] Cha, H., Kim, Y., Park, S., Yoon, T., Jung, Y., & Lee, J. (2006). Learning styles diagnosis based on user interface behaviors for the customization of learning interfaces in an intelligent tutoring system. In *Intelligent tutoring systems* (pp. 513-524): Berlin, Springer.

[3] Garcia, P., Amandi, A., Schiaffino, S., & Campo, M. (2007). Evaluating Bayesian networks' precision for detecting students' learning styles. *Computers & Education, 49*, 794–808.

[4] Carver, C. A., Jr., Howard, R. A., & Lane, W. D. (1999). Enhancing student learning through hypermedia courseware and incorporation of student learning styles.*IEEE Transactions on Education, 42*, 33-38.

[5] Gilbert, J. E., & Han, C. Y. (1999). Adapting instruction in search of 'a significant difference'.*Journal of Network and Computer Applications, 22*, 149-160.

[6] Peña, C., Marzo, J., & de la Rosa, J. (2002). Intelligent agents in a teaching and learning environment on the Web.In *Proceedings of the international conference on advanced learning technologies* (pp. 21–27): Citeseer.

[7] Papanikolaou, K., &Grigoriadou, M. (2004). Accommodating learning style characteristics in Adaptive Educational Hypermedia.In *Workshop on individual differences in adaptive hypermedia organised in AH2004, Part I* (pp. 77-86).

[8] Graf, S., &Kinshuk. (2006). An Approach for Detecting Learning Styles in Learning Management Systems. In *Advanced Learning Technologies, 2006. Sixth International Conference on* (pp. 161-163).

[9] Felder, R., & Silverman, L. (1988). Learning and teaching styles in engineering education. *Engineering education, 78*, 674-681.

[10] Özpolat, E., & Akar, G. B. (2009). Automatic detection of learning styles for an e-learning system. *Computers & Education, 53*, 355-367.

[11] Tang, T., &McCalla, G. (2005). Smart Recommendation for an Evolving E-Learning System: Architecture and Experiment. *International Journal on E-Learning, 4*, 105-129.
[12] Mor, E., &Minguillon, J. (2004). E-learning personalization based on itineraries and long-term navigational behavior. In *13th international world wide web conference* (pp. 264-265). NY.
[13] Mullier, D., Moore, D., & Hobbs, D. (2001). A Neural-Network system for Automatically Assessing Students. In P. Kommers& G. Richards (Eds.), *World conference on educational multimedia, hypermedia and telecommunications* (pp. 1366–1371): Citeseer.
[14] Drigas, A., &Vrettaros, J. (2004). An intelligent tool for building e-learning contend-material using natural language in digital libraries.*WSEAS Transactions on Information Science and Applications, 1*, 1197-1205.
[15] Marton, F. (1986). Phenomenography: A research approach to investigating different understandings of reality. *Journal of thought, 21*, 28-49.
[16] Pittenger, D. (1993). The utility of the Myers-Briggs type indicator.*Review of Educational Research, 63*, 467.
[17] Kolb, D., Boyatzis, R., &Mainemelis, C. (2001). Experiential learning theory: Previous research and new directions. *Perspectives on thinking, learning, and cognitive styles*, 227-247.
[18] Felder, R., & Brent, R. (2005). Understanding student differences.*Journal of Engineering Education, 94*, 57-72.
[19] Dunn, R., & Griggs, S. (2003). The Dunn and Dunn learning style model and its theoretical cornerstone. In R. Dunn & S. Griggs (Eds.), *Synthesis of the Dunn and Dunn learning styles model research: who, what, when, where and so what–the Dunn and Dunn learning styles model and its theoretical cornerstone* (pp. 1–6). New York: St. John's University.
[20] Felder, R. M. (1993). Reaching The Second Tier Learning And Teaching Styles In College Science Education. *Journal of College Science Teaching, 23*, 286-290.
[21] Felder, R., &Soloman, B. (1997). Index of learning styles questionnaire. In: North Carolina State University http://www.engr.ncsu.edu/learningstyles/ilsweb.html.
[22] Frank, T., Kraiss, K., &Kuhlen, T. (1998). Comparative analysis of fuzzy ART and ART-2A network clustering performance.*IEEE Transactions on Neural Networks, 9*.
[23] Grossberg, S. (1976). Adaptive pattern classification and universal recoding: I. Parallel development and coding of neural feature detectors. *Biological cybernetics, 23*, 121-134.
[24] Carpenter, G. A., &Grossberg, S. (1988). The ART of Adaptive Pattern Recognition by a Self-Organizing Neural Network.*Computer, 21*, 77-88.
[25] Carpenter, G., &Grossberg, S. (1987b). A massively parallel architecture for a self-organizing neural pattern recognition machine.*Computer vision, graphics, and image processing, 37*, 54-115.
[26] Carpenter, G., &Grossberg, S. (1987a). ART 2: Self-organization of stable category recognition codes for analog input patterns. *Applied optics, 26*, 4919-4930.
[27] Carpenter, G. A., Grossberg, S., & Rosen, D. B. (1991a). ART 2-A: An Adaptive Resonance Algorithm for Rapid Category Learning and Recognition. *Neural Networks, 4*, 493-504.
[28] Whiteley, J., Davis, J., Mehrotra, A., &Ahalt, S. (1996). Observations and problems applying ART 2 for dynamic sensor pattern interpretation. *IEEE Transactions on Systems, Man, and Cybernetics Part A: Systems and Humans, 26*, 423-437.
[29] Carpenter, G., Grossberg, S., & Rosen, D. (1991). Fuzzy ART: Fast stable learning and categorization of analog patterns by an adaptive resonance system. *Neural Networks, 4*, 759-771.
[30] Carpenter, G. A., Grossberg, S., & Reynolds, J. H. (1991). ARTMAP: Supervised real-time learning and classification of nonstationary data by a self-organizing neural network. *Neural Networks, 4*, 565-588.
[31] Carpenter, G. A., Grossberg, S., Markuzon, N., Reynolds, J. H., & Rosen, D. B. (1992). Fuzzy ARTMAP: A neural network architecture for incremental supervised learning of analog multidimensional maps. *IEEE Transactions on Neural Networks, 3*, 698-713.
[32] Carpenter, G. A., Grossberg, S., & Rosen, D. B. (1991b). Fuzzy ART: Fast stable learning and categorization of analog patterns by an adaptive resonance system. *Neural Networks, 4*, 759-771.

Gholam Ali Montazer received his B.Sc. degree in Electrical Engineering from Kh. N. Toosi University of Technology, Tehran, Iran, in 1991, his M.Sc. degree in Electrical Engineering from TarbiatModares University, Tehran, Iran, in 1994, and his Ph.D. degree in Electrical Engineering from the same university, in 1998. He is an Associate Professor of the Department of Information Engineering in TarbiatModares University, Tehran, Iran. His areas of research include Information Engineering, Knowledge Discovery, Intelligent Methods, System Modeling, E-Learning and Image Mining.

Hessam Khoshniat received his B.Sc. degree in Software Engineering from Ferdowsi University of Mashhad, Iran, in 2008. Now He is a M.Sc. student of Informan Technology Engineering at TarbiatModares University, Tehran, Iran. His research interests include Artificial Intelligence, Intelligent and Affective Tutoring Systems, Clustering and Pattern Recognition.