Expert Systems with Applications 38 (2011) 6480–6485

Contents lists available at ScienceDirect

Expert Systems with Applications

journal homepage: www.elsevier.com/locate/eswa

Data mining for adaptive learning in a TESL-based e-learning system

Ya-huei Wang [a,b,*], Hung-Chang Liao [c,b]

[a] Department of Applied Foreign Languages, Chung Shan Medical University, Taiwan
[b] Department of Medical Education, Chung Shan Medical University Hospital, No. 110, Sec. 1, Jian-Koa N. Road, Taichung 402, Taiwan
[c] Department of Health Services Administration, Chung Shan Medical University, Taiwan

A R T I C L E I N F O

Keywords:
Adaptive learning
Data mining
Neural network
e-Learning system

A B S T R A C T

This study proposes an **A**daptive **L**earning in Teaching English as a Second Language (**TESL**) for **e-learning system** (AL-TESL-e-learning system) that considers various student characteristics. This study explores the learning performance of various students using a data mining technique, an artificial neural network (ANN), as the core of AL-TESL-e-learning system. Three different levels of teaching content for vocabulary, grammar, and reading were set for adaptive learning in the AL-TESL-e-learning system. Finally, this study explores the feasibility of the proposed AL-TESL-e-learning system by comparing the results of the regular online course control group with the AL-TESL-e-learning system adaptive learning experiment group. Statistical results show that the experiment group had better learning performance than the control group; that is, the AL-TESL-e-learning system was better than a regular online course in improving student learning performance.

1. Introduction

In the conventional learning institutions of Taiwan, English teachers present the same content to all students regardless of the individual student's gender or learning characteristics. In other words, English courses are based on "static" learning material, not "dynamic" learning material (Romero, Ventura, Delgado, & Bra, 2007). This is because of the enormous costs universities must pay for education materials in Taiwan, which make it impossible to design personalized learning environments to accommodate the learning needs of individual students. In this type of learning system, if students wish to maximize their learning outcomes, they must adapt to the course content, as the course content cannot be adapted to accommodate *their* individual needs and preferences.

However, adaptive learning for individual students has recently become popular in the educational field. An adaptive learning system is a system developed to accommodate a variety of individual needs and differences. To improve student interaction and learning outcomes, several researchers have recently examined ways to develop adaptive learning for use in different courses (Arlow & Neustadt, 2001; Constantine, 2001; Gibbons, Nelson, & Richards, 2001; Larman, 2001). When educational costs are considered, e-learning is an attractive contemporary approach to achieving the goal of adaptive learning. Chen, Liu, and Chang (2006) presented a personalized web-based instruction system, based on modified item response theory, which performs personalized curriculum sequencing while simultaneously considering course difficulty and learner abilities. This approach uses the concept of learning pathways to help students learn more effectively. In addition, Chen, Hsieh, and Hsu (2007) discovered the association rules of common learning misconceptions using the testing item responses of various learner profiles for web-based learning diagnosis, and applied these association rules to promote learning performance. Tseng, Su, Hwang, Tsai, and Tsai (2008) proposed an adaptive learning system based on a modular framework that segments and transforms teaching materials into modular learning objects. Using this approach, a teacher can dynamically compose the course content according to the profiles and portfolios of individual students. Hsu (2008a) proposed a recommender teaching and learning system to help identify and address student problems and weaknesses in the English language learning process.

The data mining technique is indispensable in developing an e-learning system. Huang, Huang, and Chen (2007) used computerized adaptive testing of individual learner requirements to develop a summative examination and assessment analysis and construct a personalized e-learning system based on a genetic algorithm data mining technique. Hsu (2008b) used content-based analysis, collaborative filtering, and data mining techniques—including an association rules algorithm—to analyze students reading data and select appropriate lessons for each student. Sun, Cheng, Lin, and Wang (2008) proposed a grouping method based on data mining to establish effective groups. Their method helps teachers improve group leaning performance in e-learning. Chen and Hsu (2007) proposed a novel data mining technique consisting of tree-like

* Corresponding author at: Department of Applied Foreign Languages, Chung Shan Medical University, Taiwan.
E-mail address: huncliao@ms43.hinet.net (Y.-h. Wang).

0957-4174/$ - see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.eswa.2010.11.098

patterns that integrated a pair of items into a novel e-learning platform using their cause and effect relationships.

Based on the results of the studies above, this paper presents an adaptive leaning system that accommodates individual student needs and differences in the field of Teaching English as a Second Language (TESL). A data mining technique was used to construct the proposed e-learning system. Specifically, this paper adopts a 4-*step* approach based on an artificial neural network (ANN) core data mining technique to develop an **A**daptive **L**earning in **TESL** for **e-learning system** (AL-TESL-e-learning system). A back-propagation (BP) algorithm selected from the ANNs was used for the supervised cluster classification of student characteristics and learning performances. Different levels of teaching content for vocabulary, grammar, and reading were then set for different students with different combinations of characteristics. Finally, a control group in a regular online course and an experimental group enrolled in the AL-TESL-e-learning system were compared in the pre-test and post-test to validate the feasibility of the AL-TESL-e-learning system. The following section discusses the concept of ANNs and how to use the BP algorithm. Section 3 introduces the sample material to further verify the AL-TELS-e-learning system. Section 4 describes the 4-*step* approach for developing the AL-TELS-e-learning system. The experimental results in Section 5 confirm the proposed AL-TELS-e-learning system. Section 6 presents a summary of the paper's findings and contribution to the literature.

2. ANNs model

ANNs are composed of processing elements (nodes or neurons) and their connections. The nodes are interconnected layer-wise among themselves. Each node in each successive layer receives the inner product of synaptic weights with the outputs of the nodes in the previous layer. The operation of single node is shown in Fig. 1. ANNs have been shown to be effective for addressing complex nonlinear problems. The two types of learning networks are supervised and unsupervised respectively. For a supervised learning network, a set of training input vectors with a corresponding set of target vectors is trained to adjust weights in the ANN. For an unsupervised learning network, a set of input vectors is proposed; however, no target vectors are specified. In this study, a supervised learning network was thought to be more suitable for the classification problem. Several well-known supervised learning ANNs are the BP algorithm, learning vector quantization, and counter propagation network. The BP algorithm is used most extensively and can provide better solutions for many applications (Dayhoff, 1990; Lippmann, 1987). Therefore, the BP algorithm was selected for the current study.

A BP algorithm consists of three or more layers, including an input layer, one or more hidden layers, and an output layer. Fig. 2 illustrates a basic BP algorithm with three layers. The BP algorithm learning works on a gradient-descent algorithm (Funahashi, 1989). The BP algorithm initially receives the input vector and directly passes it into the hidden layer(s). Each element of the hidden layer(s) is used to calculate an activation value by summing up the weighted input, and the sum of the weighted input will be transformed into an activity level by using a transfer function. Each element of the output layer is then used to calculate an activation value by summing up the weighted inputs attributed to the hidden layer. Next, a transfer function is used to calculate the network output. The actual network output is then compared with the target value. The BP algorithm refers to the propagation of errors of nodes from the output layer to nodes in the hidden layer(s). These errors are used to update the network weights. The amount of weights to be added to or subtracted from the previous weight is governed by the delta rule. After the knowledge representation is determined, the BP algorithm will be trained to attempt the classification behavior. The number of hidden layers and the number of nodes in each hidden layer are determined during the training phase. In this study, a fully connected feed-forward neural network was used, and its network parameters and stopping criterion were set.

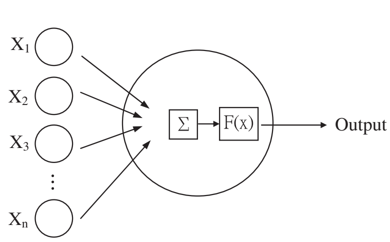

Fig. 1. Node operation.

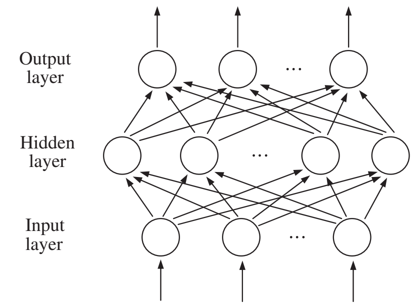

Fig. 2. A BP neural network.

To be able to attempt the classification behavior, a learning rule is used in the BP algorithm. In the case of a multi-layer perception, this rule should also be able to adapt the weights of all connections in order to model a nonlinear function. The learning rule used most frequently for this purpose is the BP rule. It acts through the following two steps. First, the generalized difference $D_i^*(t)$ is calculated by

$$D_i^*(t) = (A_i^*(t) - A_i(t)) * A_i(t) * (1 - A_i(t)), \tag{1}$$

where $A_i^*(t)$ is the desired activation of output unit i, and $A_i(t)$ is the generated activation of this unit. In order to obtain the generalized difference $D_i^*(t)$, the calculated difference $(A_i^*(t) - A_i(t))$ is multiplied by the simplified derivative of the activation function $A_i(t) * (1 - A_i(t))$. Second, the generalized differences of the units in the output layer are propagated back through the weighted connections to the units of the hidden layer(s). The generalized difference collected from a hidden unit is multiplied by the simplified derivative of the unit's activation function in order to obtain the generalized difference of the hidden unit

$$D_j^*(t) = \sum_{i=1}^{n} (W_{ij}(t) * D_i^*(t)) * A_j(t) * (1 - A_j(t)). \tag{2}$$

Using the generalized difference $D_i^*(t)$, the weights are adjusted by

$$W_{ij}(t+1) = W_{ij}(t) + C * D_i^*(t) * A_j(t). \tag{3}$$

The adaptation size of the weight $W_{ij}(t)$ of the connection used to send information from unit j to unit i is influenced by the existing weight $W_{ij}(t)$, the learning rate C, the generalized difference $D_i^*(t)$, and the actual activation $A_j(t)$ of unit j. To reduce the probability of weight change oscillation, a weight momentum term is added to adjust the weight. The weight momentum term is constructed by the previous adjustment of the weight $D * W_{ij}(t)$ and a constant value B, so