

E-Learning Optimization Using Supervised Artificial Neural-Network

Mohamed Sayed^{1,2}, Faris Baker¹

¹Faculty of Computer Studies, Arab Open University, Kuwait City, Kuwait

²On Leave from Faculty of Engineering, Alexandria University, Alexandria, Egypt

Email: msayed@aou.edu.kw, fbaker@aou.edu.kw

Received 20 December 2014; accepted 8 January 2015; published 21 January 2015

Copyright © 2015 by authors and Scientific Research Publishing Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

Improving learning outcome has always been an important motivating factor in educational inquiry. In a blended learning environment where e-learning and traditional face to face class tutoring are combined, there are opportunities to explore the role of technology in improving student's grades. A student's performance is impacted by many factors such as engagement, self-regulation, peer interaction, tutor's experience and tutors' time involvement with students. Furthermore, e-course design factors such as providing personalized learning are an urgent requirement for improved learning process. In this paper, an artificial neural network model is introduced as a type of supervised learning, meaning that the network is provided with example input parameters of learning and the desired optimized and correct output for that input. We also describe, by utilizing e-learning interactions and social analytics how to use artificial neural network to produce a converging mathematical model. Then students' performance can be efficiently predicted and so the danger of failing in an enrolled e-course should be reduced.

Keywords

Artificial Neural Networks, E-Learning, Prediction Models, Supervised Learning

1. Introduction

Education is imperative for every nation for it improves the life of individuals by training them with skills and knowledge that let them cope with life challenges. Technological developments today influence every aspect of life including education. Technology provides speed and convenience for people and hence becomes a vital instrument in the educational process [1]. In blended learning model [2] where face to face learning is combined with technology, students and tutors as well as other stakeholders use computers and internet to communicate

and collaborate. The widespread of social networks has inspired some educational institutions to investigate the impact of this paradigm on the learning outcome. The use of internet provides opportunities to analyze the electronic activities performed for capturing patterns, trends and intelligence. In blended learning students watch interactive lectures, take quizzes through a learning management system (LMS) and hence prepare themselves for the coming face to face class with their tutors. In this way, tutors spare their time for discussion and solving student's difficulties. The LMS provides not only e-courses for students, but also a platform to communicate with peers and tutors. Anytime a student or a tutor login into the LMS, a digital record is stored representing the activities performed. In the end of each semester, a large amount of data will be generated regarding every course offered in the university. The data generated illustrate the successes, and the difficulties encountered in the learning process.

Educational theorists [3]-[5] have identified some parameters that contribute to the learning successes. These factors include student engagement, student self-regulation, student interaction with his/her peers, tutors total experience and tutors time involvement with their students. In addition, the support of the university administration contributes to a higher achievement of students as well as tutors. Moreover, a meaningful engagement which includes a high quality of discourse is imperative in learning and gives an impetus for sharing information. All these parameters have to be investigated in the large data generated during the semester and analyzed to improve and correct the process of learning [6] [7].

Inspired by advances in social networks analytics, the document analysis concept is carried out through the study of engagement in the e-learning during a semester identifying several variables that describe the student engagement. The network of co-occurrences between different variables, collected on a specific set, allows the quantitative study of the structure of contents, in terms of the nature and intensity level of correlations or inter-connections. The sub-domains are placed using the structural equivalence techniques by grouping variables at different stages. A scientific field is characterized by a group of variables, which signify its concepts, operations and methodologies. The structure described by the frequency of co-occurrences of conceptual variables exposes the important relationships across these variables. These analyses of co-occurrences of variables give us the authority to comprehend the static and dynamic sides of the room in which we can relate and place their work in a hierarchy of scientific research concepts. This technique assorted as co-variable analysis, provides a direct quantitative manner of linking the conceptual contents.

According to Edelman [8], there are five prevalent kinds of information: sequences, association, classifications, clusters and prediction. Likewise, the general purpose of neural networks is to provide powerful solutions to associations, classifications, clusters and predictions problems. Moreover, neural networks possess the destiny to impart from experience in order to change for the better by improving their performance and adapting themselves. The neural networks are also able to transact with deprecatory information (incomplete or noisy data) and can be very efficient, in particular where it is not possible to designate the patterns or steps that take part to the resolution of a problem. In many cases neural network techniques tend to be difficult to understand and there is no clear advantage over more conventional techniques. However, there are features of neural networks which distinguish them from other techniques. The fact that the network can be trained to perform a task, and the ability of the network to generalize from the training data enables it to carry out successfully when presented with a previously unseen problem.

In the present research, a normalized co-variable matrix from 56 most-used categorized variables (features or predictors) is used to study the contribution to learning process. This matrix is split using its mean density to conclude the correlation matrix and build the network map. In order to achieve higher levels of computational capacity, an exceedingly complex structure of neural networks is required. We use a multilayer perceptron neural network which maps a set of input variables onto a set of output data. It consists of multiple layers of nodes in a directed graph at which each layer is connected to the next one. It is one of the most popular and successful techniques for many subjects such as content analysis, pattern recognition and document image analysis. It is also a potent technique to solve many real world problems such as predicting time-series in monetary world, identifying clusters of valuable customers, and diagnosing medical conditions and fraud detections, (see for instance [9]-[12]).

The multilayer perceptron neural network has not been applied comprehensively, to the best of our knowledge, to e-learning optimization using supervised and unsupervised learning. The questions which arise then are whether the neural network technique is indeed appropriate to such problem, whether the architecture used to implement the technique reduces its effectiveness or complements it, and whether the technique produces a par-

ticular system that attaches to the problem.

2. Theoretical Framework and Neural Network

The aim of regression methods is to provide a definite model which can be helpful in deriving a specific group that one of the database objects belongs to based on its features. One of the usual functions of regression methods includes determining future activities of student engagement so that the institute could alter the e-learning strategy [13].

Data mining is an automatic analysis technique in large data sets whose purpose is to extract unobserved correlations (or dependencies) between the data stored in data ware houses or other relational database schemes. The end-user may even not be aware of these data correlations and dependencies, although the knowledge derived from extracting them may turn out to be exceedingly beneficial. Data mining techniques [14] may be divided into several basic groups, each sharing with specific purposes and complications. The k -nearest neighbors (k -NN) algorithm is one of the most popular and non-parametric method used in classification and regression. Its purpose is to find k -nearest neighbors of the discipline (using some predefined metrics) and ascribe it to a certain class that is predominant in all successfully found subjects. A drawback of skewed class distribution renders this method of classification less acceptable. Another commonly used technique which accepts the vantage of being comparatively unsophisticated and efficient is the naïve (or simple) discriminate analysis. There are other techniques that require a learning set form a numerous groups such as artificial neural networks, simple and oblique decision trees, and support vector machines (SVM) methods. All of these techniques are founded on a similar principle that consists of choosing a structure (for example: multi layered perception for neural networks, leaves representing class labels and branches representing conjunction of features for decision tree, and core function for the SVM method). Some of these methods consist of putting the best parameters that permit to minimize erroneous classifications on the given learning set (for instance: using the error back propagation method or optimization methods).

The artificial neural network (ANN) is a parallel and iterative method made up of simple processing units called neutrons. While a multilayer neural network is a web of simple neurons called perceptron. The principle concept of a single perceptron was introduced by Frank Rosenblatt in 1958. A multilayer perceptron neural network (MLP) is a perceptron-type network which distinguishes itself from the single-layer network by having one or more intermediate layers. Backward propagation of errors (or simply back propagation), which has been used since the 1980s to adjust the weights, is a widespread process of training artificial neural networks. It is usually used in conjunction with an optimization method such as gradient descent. In an attempt to minimize the loss function; on each of training iteration, the current gradient of a loss function with respects to all the calculated weights in the network is evaluated and then the gradient is fed to the optimization method which employs it to update the weights. In this study we selected the standard Levenber-Marquardt Algorithm (LMA) [15], a curve-fitting algorithm that minimizes the sum of the squares of errors between the data points and the parameterized loss function. In some few cases, the process of training (or learning) could lead to over (or under) training phenomena and hence one may prove that it is time-consuming.

3. Data Set Description

In order to study the variables that contribute to the learning process and to the educational outcome we propose to consider specific variables during the semester of engagement relevant to students, to the peers, to the tutor and to the university administration. The 56 variables, see **Table 1**, are categorized according to its characteristics and ownerships. Moreover, we have to decide the most suitable machine learning algorithm for the selected features regression.

Data is contextual, the sequences and the environment of the data has also to be accounted for. For example, an activity performed by a student who has a limited background is not the same as the one who has the entire necessary prerequisite. After data acquisition from the online activities performed by all participants, filtering has to be applied in order to remove irrelevant data. Once all the data are clean and relevant, a second stage of features extraction, clustering and classifications is applied in order to extract knowledge from the data. As mentioned above, there are many methods for features extraction and regression such as artificial neural network, decision tree, Markov model, Bayesian probability, principle component analysis, support vector machine, and regression analysis [16]. In this paper, the explored variables are configured in a back propagation neural network

Table 1. Categorized data set description.

Class	Variable	Full name	Class	Variable	Full name	
Student online engagement	Var. 1	Login into the e-class.	Student content analysis	Var. 30	Course keywords used.	
	Var. 2	Participate in the e-learning forums.		Var. 31	Question context.	
	Var. 3	Starts a new thread in the forums.		Var. 32	Answer context.	
	Var. 4	Reads a thread from a classmate.		Var. 33	Comment context.	
	Var. 5	Votes on a post reply as “LIKE”.		Var. 34	Disputation comment.	
	Var. 6	Votes on a post reply as “DISLIKE”.		Var. 35	Cumulative comment.	
	Var. 7	Receives a “LIKE” reply.		Var. 36	Exploratory comment.	
	Var. 8	Receives a “DISLIKE” reply.		Var. 37	Effective interaction.	
	Var. 9	Starts a thread that a classmate votes up.		Var. 38	Ineffective interaction.	
	Var. 10	Starts a thread that a classmate votes down.		Tutors online engagement	Var. 39	Answer a question.
	Var. 11	Enters the online quizzes.			Var. 40	Give a comment.
	Var. 12	Solves the online questions.			Var. 41	Start a thread.
	Var. 13	Answers the online question at first instance.		E-course design	Var. 42	Number of questions provided.
Var. 14	Enjoyment.	Var. 43	Number of hours per week required from a student.			
Var. 15	Anxiety.	Var. 44	Personalized (relevant to student background).			
Student self-regulation	Var. 16	Boredom.	University support	Var. 45	System downtime.	
	Var. 17	Hopelessness.		Var. 46	Low participated student identification.	
	Var. 18	Self-efficacy.		Var. 47	High participated student identification.	
	Var. 19	Effort.		Var. 48	System bandwidth.	
	Var. 20	Ambition.		Var. 49	Emails send to the student.	
	Var. 21	Goal oriented.		Var. 50	Emails send to the tutor.	
	Var. 22	Self-organized.		Student information	Var. 51	Year of admission.
Var. 23	GPA before the e-course.	Var. 52	Semester of admission.			
Var. 24	Average grades of prerequisite courses for the e-course.	Var. 53	Total credit hours.			
Var. 25	High School GPA.	Var. 54	Number of warnings.			
Var. 26	Family education.	Var. 55	Student status.			
Var. 27	Number of family members.	Var. 56	Over all GPA.			
Var. 28	Family income.					
Var. 29	Residency location.					

algorithm that will investigate their impact on the educational outcome.

The identified values of the variables will be used to train the neural network. For more information about the variables the reader is referred to [3] [11] [17]-[20].

4. Backward Propagation Algorithm

Multilayer neural network is a particular type of network which consists of a group of sensory units. These units are observed as cascading layers; an input layer, one (or more) intermediate-hidden layers and an output layer of neurons. The neural network is completely connected such that all neurons of each layer are connected to all neurons in the preceding layer. At the beginning of the backward propagation process we should consider how

many hidden layers are required. The computational complexity can be seen by the number of single-layer networks combined into this multilayer network. In this multilayer structure, the input nodes pass the information into the units in the first hidden layer then the outputs from the first hidden layer are passed into the next layer, and so on. It is worth noting that the network is a supervised learning, *i.e.*, both the inputs and the outputs should be provided. The network processes the inputs and compares its resulting outputs against the desired corresponding outputs. Errors are then calculated, giving rise to the system to regularize the weights and control the network. This process takes place over and over as the weights are continually adjusted.

The back propagation algorithm of multilayer neural network is summarized in forward and backward stages. In the forward stage, the signal that transfers out of the network (through the network layers) is calculated as follows:

$$s_j^l = \sum_{i=0}^{n^{l-1}} w_{ji}^l y_i^{l-1}$$

where $l = 0, 1, \dots, L$ are network layers indexes, $l = 0$ and $l = L$ represent the input and output layers respectively, n^l corresponds to the number of neurons in the layer l and $j = 0, 1, \dots, n^{l-1}$. Here y_i^{l-1} is the output function corresponding to the neuron i in the previous layer $l-1$ and w_{ji}^l is the weight of the neuron j to the neuron i in the layer l . In addition to the variable weight values, an extra input that represents bias is added, *i.e.*, for $i = 0$, $y_0^{l-1} = 1$ and w_{j0}^l depict the bias that is applied to neuron j in the layer l . The output of neuron j in the layer l is given by the neuron activation function of j :

$$y_j^l = f_j(s_j^l).$$

The error at the output layer is

$$e_j^l = y_j^l - O.$$

It represents the difference between the target output for an input pattern and the network response. It is used to calculate the errors at the intermediate layers. This is done sequentially until the error at the very first hidden layer is computed. After computing the error for each unit, whether it is at a hidden unit or at an output unit, the network then fine-tunes its connection weights by performing the backward. The general idea is to use the gradient descent to update the weights so that the square error between network output values and the target output values are reduced. The backward can be performed as follows:

$$z_j^{l-1} = f'(s_j^{l-1}) \sum_{i=0}^{n^l} w_{ji}^l z_i^l$$

where $l = L, L-1, \dots, 2$, $j = 0, 1, \dots, n^{l-1}$ and

$$z_j^L = f'(s_j^L) e_j^L.$$

The learning rate which is a typically a small value between 0 and 1 controls the size of weight modulations. Here the derivative $f'(x)$ governs the weight adjustments, depending on the actual output $f(x)$.

The major problem in training a neural network is deciding when to stop the training. The algorithm brings to an end when the network reaches minimum errors which can be calculated by the mean square errors (MSE) between the network output values and the desired output values. The number of training and testing iterations can also be used as stopping criteria.

5. Training the Network and Discussion

Training the neural network to emerge the right output for a given input is a computational iterative procedure. The evaluated root mean square error of the neural network output (on each training iteration) and the way which the error changes with training iteration are utilized to determine the convergence of the training. The challenge is to determine which indicators and input data could be practiced, and to amass enough training data to improve the network appropriately. Many factors interact with each other to generate the observed data. These factors are organized into multiple layers, representing multiple abstractions, weights and biases. By using various numbers of layers and neurons, different levels of abstractions spawn with different features. It is possible to train a neural network to perform a particular function by adjusting the weights of the connections between the neurons. Errors are propagated backward through the network to control weight adjustments. Network

layers are trained when errors fall below a threshold.

The process of training the neural network is summarized as follows: input data is continuously applied, actual outputs are calculated, and weights are adjusted such that the application of inputs produce desired outputs (as close as possible). Weights should converge to some value after many rounds of training and differences between desired and actual outputs are minimized.

In our experiment, we observed 1879 students (in one semester) using student information criteria, mentioned **Table 1**, as a sample for our educational inquiry. We used 70% of the data (1315 samples) for training, 15% of the data (282 samples) for validation, and 15% of the data (282 samples) for testing. The training data were represented to neural network, and the network was adjusted according to its error. The validation was used to measure network generalization, and to halt training when generalization stops improving as indicated by an increase in the mean square error of the validation samples. The testing data have no effect on the training; they only provide performance measurement during and after training.

The first run of the algorithm using 50 hidden neurons produced the result in **Table 2**. The Mean Square Error measured the average squared differences between outputs and targets while regression values R calculated the correlation between outputs and targets. An R value of 1 means a close relationship between the calculated output and the target while a value of 0 means no relationship. Plotting the histogram of error values, as presented in **Figure 1**, showed the majority of the points fall between -0.6 and $+0.7$ and mostly around 0.04 values. This indicates that the training algorithm has only few outliers and generally of a good prediction outcome. Examining the regression values R (**Figure 2**), demonstrated the validation scatter diagram, which plots the actual data from validation portion against the neural network estimate, gave a straight-line graph (independent and objective validation metric) with a slope extremely close to one and an intercept close to zero. The validation is comprehensive as the data is picked out at random over all the available data set. The training data were represented to the neural network algorithm, and the network was adjusted to minimize its error, the validation measured network generalization, and halted training when generalization stopped improving. The mean square errors stopped decreasing after iteration (epoch) 12 (**Figure 3**). During the training of the algorithm, an increase in the mean square error will stop the training of the validation samples. Bearing in mind training multiple times will

Table 2. Results of the algorithm with 50 hidden neurons.

	Samples	MSE	R
Training	1315	1.81246e-1	9.20899e-1
Validation	282	3.33569e-1	8.64064e-1
Testing	282	2.21908e-1	9.09618e-1

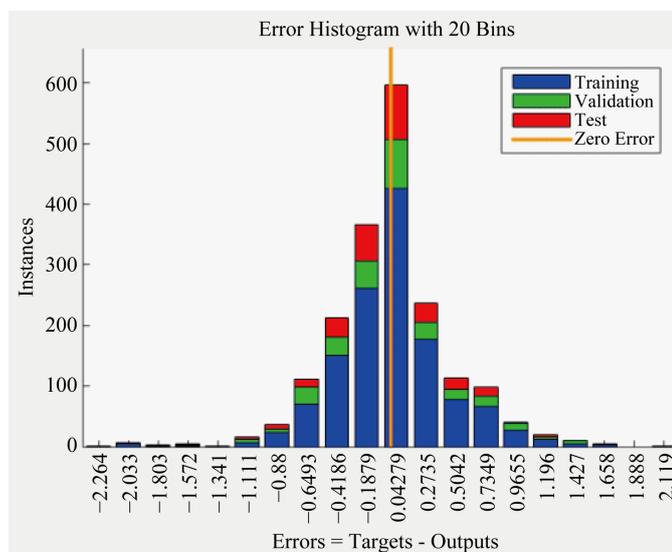


Figure 1. Plot of the error histogram.

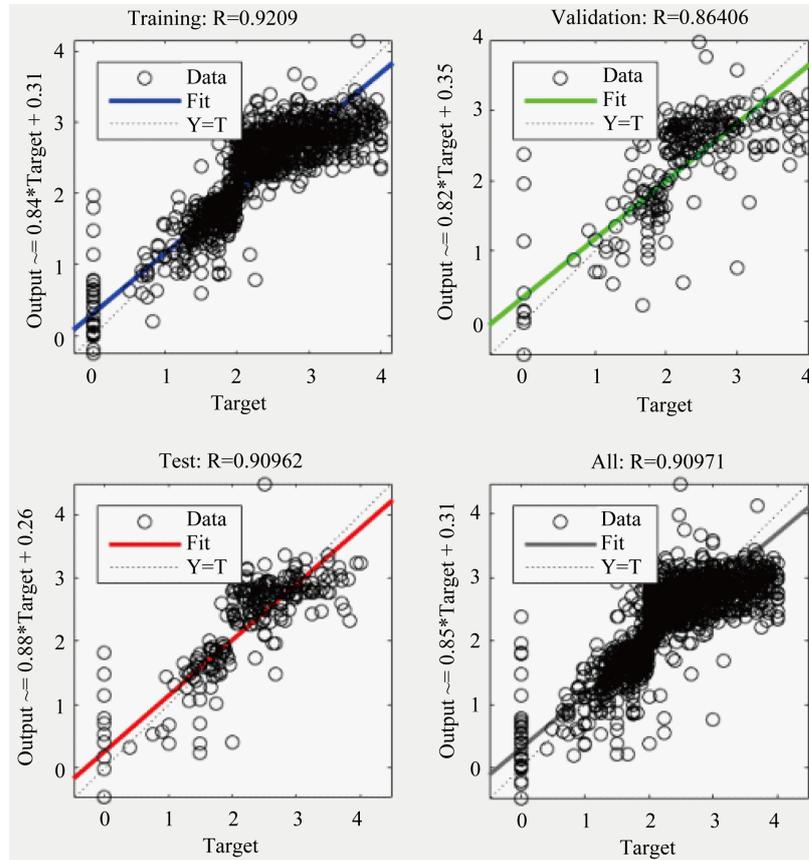


Figure 2. Plot of the regression values (R).

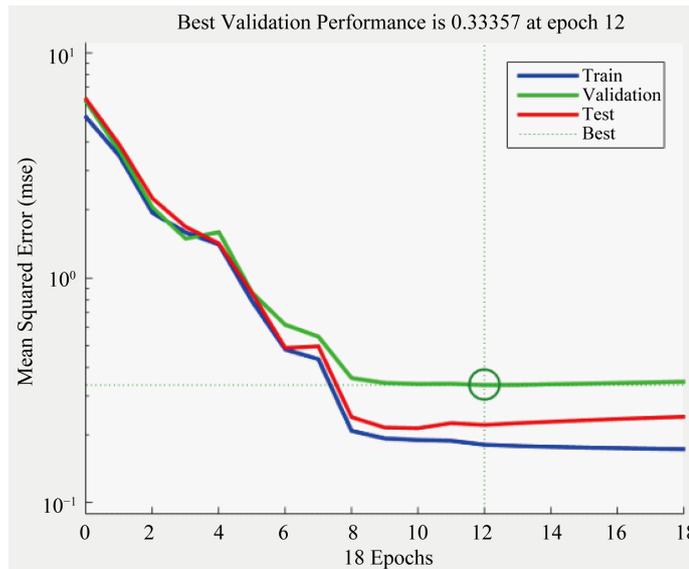


Figure 3. Plot of the performance.

generate different results due to different initial conditions and sampling.

After iteration 12, the gradient decent calculated by back propagation algorithm was not increasing and in a further 6 validation iterations, hence stopped at epoch 18 with value 0.035909 (Figure 4). At iteration 12, the validation stopped after 6 trails because the error between the outputs and the target values was not decreasing

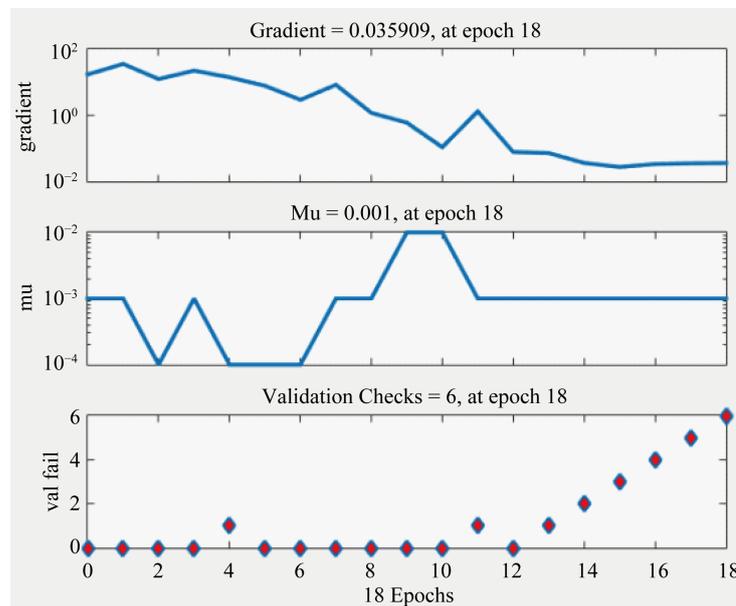


Figure 4. Plot of the training state.

further as well as Mu values. The neural network became ready and trained to perform the desired function which is to predict the Grade Point Average (GPA) for future students, provided predictors are available.

6. Conclusions

Neural networks learn from examples and capture functional relationships between variables in the data even if the underlying relationships are nonlinear or unknown. Even though neural networks are not perspicuous in their expectation, they can outperform all other methods of association, classifications, clusters and prediction of supervised and unsupervised learning as proved with their high performance prediction for non-linear systems. Furthermore, the training algorithm may change depending on the neural network structure, unless the nearly common training algorithm used when designing networks is the back-propagation algorithm. The major problem in training a neural network is deciding when to finish operations as well as the overtraining phenomena which occur when the system memorizes patterns and thus lacks the power to extrapolate.

Due to our research constraint in our experiment, we selected only subset predictors in our training algorithm for students GPA; however, in future research it can be extended to all other variables that have not been selected and hence improve the performance outcome. The nature and the causes of the correlations between the predictors have to be explored. Furthermore, there are opportunities to experiment with other learning algorithms and contrast with neural network.

References

- [1] Soller, A. and Lesgold, A. (2003) A Computational Approach to Analyzing Online Knowledge Sharing Interaction. *Proceedings of Artificial Intelligence in Education*, Sydney, 20-24 July 2003, 253-260.
- [2] Sayed, M. and Baker, F. (2014) Blended Learning Barriers: An Investigation, Exposition and Solutions. *Journal of Education and Practice*, **5**, 81-85.
- [3] Vohs, K.D. and Baumeister, R.F. (2011) *Handbook of Self-Regulation: Research, Theory, and Applications*. 2nd Edition, Guilford Press, New York.
- [4] Zhao, C.M. and Kuh, G.D. (2004) Adding Value: Learning Communities and Student Engagement. *Research in Higher Education*, **45**, 115-138. <http://dx.doi.org/10.1023/B:RIHE.0000015692.88534.de>
- [5] Zimmerman, B.J. (2008) Investigating Self-Regulation and Motivation: Historical Background Methodological Developments, and Future Prospects. *American Educational Research Journal*, **45**, 166-183. <http://dx.doi.org/10.3102/0002831207312909>
- [6] Ferguson, R. and Shum, S.B. (2012) Social Learning Analytics: Five Approaches. *Proceedings of the 2nd International*

- Conference on Learning Analytics and Knowledge*, Vancouver, 29 April-2 May 2012, ACM Press, New York. <http://dx.doi.org/10.1145/2330601.2330616>
- [7] Yoo, J. and Kim, J. (2014) Can Online Discussion Participation Predict Group Project Performance? Investigating the Roles of Linguistic Features and Participation Patterns. *International Journal of Artificial Intelligence in Education*, **24**, 8-32. <http://dx.doi.org/10.1007/s40593-013-0010-8>
- [8] Edelstein, H.A. (1999) Introduction to Data Mining and Knowledge Discovery. 3rd Edition, Crows Corporation, Potomac.
- [9] Chang, C.Y., Wang, H.J. and Li, C.F. (2010) Image Content Analysis Using Modular RBF Neural Network. *Journal of Computers*, **21**, 39-52.
- [10] Haykin, S. (1998) Neural Network: A Comprehensive Foundation. 2nd Edition, Prentice Hall, New York.
- [11] Srimani, P.K. and Kamath, A.S. (2012) Neural Networks Approach for the Performance Analysis of Learning Model—A Case Study. *International Journal of Current Research*, **4**, 236-239.
- [12] Srimani, P.K. and Kamath, A.S. (2012) Data Mining Techniques for the Performance Analysis of a Learning Model—A Case Study. *International Journal of Computer Applications*, **53**, 36-42.
- [13] Romero, C. and Ventura, S. (2007) Educational Data Mining: A Survey from 1995 to 2005. *Expert Systems with Applications*, **33**, 135-146. <http://dx.doi.org/10.1016/j.eswa.2006.04.005>
- [14] Bigus, J.P. (1996) Data Mining with Neural Networks: Solving Business Problems from Application Development to Decision Support. McGraw Hill, New York.
- [15] Liu, H. (2010) On the Levenberg-Marquardt Training Method for Feed-Forward Neural Networks. *Proceedings of the 6th International Conference on Natural Computation*, Yantai, 456-460.
- [16] Guyon, I. and Elisseeff, A. (2003) An Introduction to Variable and Feature Selection. *The Journal of Machine Learning Research*, **3**, 1157-1182.
- [17] Mercer, N. (2004) Sociocultural Discourse Analysis: Analyzing Classroom Talk as a Social Mode of Thinking. *Journal of Applied Linguistics*, **1**, 137-168. <http://dx.doi.org/10.1558/japl.2004.1.2.137>
- [18] Sayed, M. (2013) Blended Learning Environment: The Effectiveness in Developing Concepts and Thinking Skills. *Journal of Education and Practice*, **4**, 12-17.
- [19] Tempelaar, D.T., Niculescu, A., Rienties, B., Gijsselaers, W.H. and Giesbers, B. (2012) How Achievement Emotions Impact Students' Decisions for Online Learning, and What Precedes Those Emotions. *The Internet and Higher Education*, **15**, 161-169. <http://dx.doi.org/10.1016/j.iheduc.2011.10.003>
- [20] Winne, P.H. and Baker, R.S. (2013) The Potentials of Educational Data Mining for Researching Metacognition, Motivation and Self-Regulated Learning. *Journal of Educational Data Mining*, **5**, 1-8.

Scientific Research Publishing (SCIRP) is one of the largest Open Access journal publishers. It is currently publishing more than 200 open access, online, peer-reviewed journals covering a wide range of academic disciplines. SCIRP serves the worldwide academic communities and contributes to the progress and application of science with its publication.

Other selected journals from SCIRP are listed as below. Submit your manuscript to us via either submit@scirp.org or **Online Submission Portal**.

