International Journal of Engineering Applied Sciences and Technology, 2016
Vol. 1, Issue 9, ISSN No. 2455-2143, Pages 6-10
Published Online July – August 2016 in IJEAST (http://www.ijeast.com)

A NOVEL METHOD FOR TEACHING QUALITY ASSESSMENT BASED ON DEEP LEARNING

Gang Zhang[1], Yonghui Huang[1], Ling Zhong[1], Xuyu Sun[2]
[1] School of Automation, Guangdong University of Technology, Guangzhou, China, 510006
[2] School of Innovation and Entrepreneurship Education, Guangdong Peizheng College, Guangzhou, China, 510830

Abstract— **Teaching quality assessment is an effective way to improve the teaching quality of a university. A common way for teaching quality assessment is to make a survey with a set of well-designed questionnaires and score them. However, it cannot accurately provide good assessment since simply scoring cannot capture the potential concepts in teaching activities. In this paper we provide a novel assessment method based on deep learning, a hot topic in machine learning community. We design stacked auto-encoders to transform the original questionnaire options into high level concept vectors and then place a learner for regression in order to get set of scores for assessment. The model can also incorporate various information gathered from teaching activities. We evaluate the proposed method in our university and the results show that it is superior to previous ones.**

Keywords— **teaching quality assessment, deep learning, stacked auto-encoders, questionnaire scoring, ensemble learning, regression**

I. INTRODUCTION

Teaching quality assessment is an effective way to improve the teaching quality of a university. Currently in universities the number of departments is increasing dramatically [1]. Teaching activities of different departments are significantly different in either form of organization or styles [2, 3]. Meanwhile, the standards for evaluating student's learning performance are different in many aspects. It poses a great challenge to evaluating the teaching quality, either for the whole department or individual teachers [4, 5].

However, teaching quality assessment is an important information source for administrators and decision makers in a university. An accuracy and close-to-the-ground-truth assessment report is appreciate for a university [6]. Commonly, teaching quality assessment is performed in a questionnaire-based manner. Students are required filled a set of questionnaires when a semester is over. And a scoring algorithm is applied to the filled questionnaires and finally a score or a vector of scores is obtained. In some universities, forms filled by some teaching supervisors (experienced teachers) are used as a complement information source.

Scoring-based algorithms have significant shortcomings. Firstly, it is controlled by a small set of parameters which are manually set according to some expert knowledge and experiences. Secondly, a scoring-based algorithm is usually a shallow model which cannot capture potential concepts in teaching data. To overcome these shortcomings, in this paper we propose a teaching quality assessment method based on deep learning. The method uses stacked auto-encoders [7] as an encoding network which encodes teaching data gathered from different information sources into high level concepts. Stacked auto-encoders (SAE) is a deep learning model composing of several stacking one by one encoders. An encoder layer imposes a nonlinear transformation on the input vector and generates an output vector with different dimension [8]. A constraint is that the output vector can be re-transformed to the input vector through the same layer having least information loss. This mechanism forces the model revealing potential concepts while keeping as much as possible original information. We place a regression learner at the output layer to obtain final scores.

We briefly review some recently successful work on this topic. Bengio [9] gave a comprehensive review of current deep learning research and future directions. Wang et al. [7] proposed a deep learning model based on auto-encoder for dimensionality reduction and applied the proposed model to multimedia data annotation. Zhong and Tezuka [10] proposed a parametric learning method to determine the optimal parameters of a convolution neural network (CNN). Their method provides an effective way to determine the model parameters of a CNN. Zhou et al. [11] proposed a fuzzy deep belief network (DBN) for semi-supervised classification. Their method incorporates unlabeled data into a DBN model and improves the generalization ability of the model with the distribution evaluation of both labeled and unlabeled data. Li et al. [12] proposed a CNN based deep learning model for image purification. Meanwhile, deep learning methods have been successfully used in various application background, such as music classification, natural language processing, voice recognition and big data analysis. We believe that deep learning model can be applied to teaching quality assessment. The reason is twofold. On the one hand, teaching quality is a subjective concept which cannot easily be evaluated by some formulas or a set of static criteria. A model having the ability to evaluate complex functions is desired for the assessment. On the other hand, to model the subjective scores of teaching supervisors, a network model with several hidden layers is the best choice. Since it has similar structure as a human brain and regards the computation behavior as the activation of a chain of nodes (cells). Due to the biological background of deep learning, it has been regarded as the most powerful method to model complex concepts.

The rest of the paper is organized as follows. Section II presents the deep learning based assessment model. Section III reports the evaluation results of the proposed model followed by some discussion. And Section IV concludes the paper.

II. MAIN MODEL

A. Deep model

We first present the basic theory of SAE and then describe the design of the encoding network used in this study. An auto-encoder is a network model with an input layer, a hidden layer and an output layer. Each pair of layers are fully connected with each other and each connection has a weight ranging from 0 to 1. The input layer and output layer of an auto-encoder have different dimensions. Figure 1 shows an example of an auto-encoder.

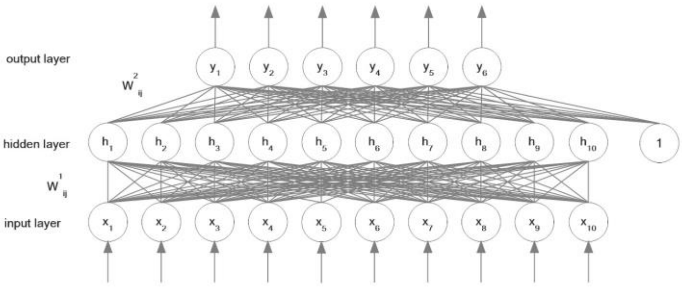


Fig. 1. An example of auto-encoder

In Fig. 1, we show an auto-encoder with a 10-node input layer, a 11-node hidden layer and a 6-node output layer. In another word, the model performs a dimensionality reduction from input to output. There are weight matrices W^1 and W^2 storing the connection weight between nodes belonging to the adjacent layers. Note that there are full connections between input/hidden and hidden/output layers. The node 1 is a bias term. Eq. (1) shows the action of a node in hidden layer performs.

$$h_j = sig(\sum_{i=1}^{10} W^1_{ij} x_i + W^1_{0j}) \quad (1)$$

In Eq. (1), the function sig stands for a standard sigmoid function. W^1_{0j} is the weight of the connection between the bias term (node 1) and the jth node in the hidden layer. The action of a node in the output layer is shown in Eq. (2).

$$y_j = \sum_{i=1}^{10} W^2_{ij} h_i + W^2_{0j} \quad (2)$$

An auto-encoder can be trained by placing a classifier on the top of the output layer. The difference between the model output and the ground truth label could be evaluated and back-propagates through the network just as the famous back propagation (BP) algorithm in training a traditional neural network.

To obtain powerful representation ability and improve the possibility to extract high level concepts, several auto-encoders can be stacked together. In this case, the output of a hidden layer of an auto-encoder does not directly feed into an output layer, but feeds into a hidden layer of another auto-encoder (often having different number of nodes in the hidden layer). The stacked hidden layers may use different activation functions other than sigmoid function. Table -1 shows the design of a stacked auto-encoders with 5 hidden layers.

Table -1 The design of a stacked auto-encoders

	nodes	**type**	**number of connections**
input	**158**	**N/A**	**158**
h1	**180**	**Full**	**180*158**
h2	**250**	**Full**	**250*180**
h3	**150**	**Full**	**150*250**
h4	**100**	**Full**	**100*150**
h5	**64**	**Full**	**64*100**
output	**64**	**Full**	**64*64**

In Table -1, the first column stands for the layers in the stacked auto-encoders. The input layer has 158 nodes because each record in our teaching quality assessment dataset has 158 elements. The column **type** stands for the connection type between the current and the next layer. The value **Full** means it is full connection. The column **number of connections** stands for how many edges between the current layer and the next layer. Note that each edge is associated with a real-value weight.

For the output layer, a 1-of-k coding function can be used for multiple-class classification tasks. In this study we use a softmax function at the top of the model. A softmax function is defined as Eq. (3).

$$softmax(y_i) = \frac{\exp y_i}{\sum_i \exp y_i} \quad (3)$$

The training target of an auto-encoder is to find a different data representation of the input vector with least information loss, i.e. the output vector can be restored to the original input through the inversed network and the classifier can work well with the encoded vector.

B. Scoring with ensemble regressors

Since our goal is to give real-value scores measuring the teaching quality, a regressor should be designed and placed at the output layer of the model. To construct a model with good generalization ability, we propose to design the regressor in an ensemble manner. We use three types of radial basis functions as our base regressors, Gauss function, Reflected Sigmoidal function and Inverse Multi-quadrics function, which are defined as Eq. (4)-(6).

$$\phi_1(r) = \exp(-\frac{r^2}{2\delta^2}) \tag{4}$$

$$\phi_2(r) = \frac{1}{1+\exp(\frac{r^2}{\delta^2})} \tag{5}$$

$$\phi_3(r) = \frac{1}{\sqrt{r^2+\delta^2}} \tag{6}$$

In Eq. (4)-(6), r stands for input variable and δ stands for the width parameter of the radial basic function. Since in our case the input is vector, we use the dot product for the square, i.e. $r^2 = r^T \cdot r$.

As mentioned in the literatures on ensemble learning, two criteria control the quality of ensemble learner. The first is accuracy and the second is diversity. Accuracy means that base learners perform well in the training dataset or validation dataset. Since in our study all base learners are well trained with all the training data, their accuracy rates can be guaranteed. For the criteria of diversity, it means that there should be some significant difference between base learners. To achieve enough large diversity of base learners, we impose a Gaussian distribution on the model parameter δ and draw N times from the distribution to get N different base learners. The distribution is designed as follows. The mean is set to the average width (distance) of all data in the dataset, as shown in Eq. (7):

$$u = \frac{1}{N \times N} \times \sum_{i=1}^{N}\sum_{j=1}^{N} d(x_i, x_j) \tag{7}$$

In Eq. (7), N stands for the size of the whole dataset, $d(\cdot,\cdot)$ is the Euclidean function. The variance σ is defined as 5 times of the variance of all $d(x_i, x_j)$ so as to provide high diversity.

For model training, we apply a layer-wise training strategy which is widely used in deep learning study. For each auto-encoder, the whole training dataset is used for training. The training begins from the layer h1 with raw inputs. For layers h2 to h5, the inputs are the outputs of the trained previous layers so as to meet the requirement of input dimensionality.

III. Experiment and Result

We evaluate the proposed method on a teaching quality dataset gathered from our university. The dataset contains 3 basic forms and totally there are 5720 records, each of which corresponds to a student of grade 2 / 3. Table -2 shows the details of the main form for normal courses.

Table -2 Basic questions for normal course

Categories	No.	Question
Teaching attitude	1	be late/leave early
	2	dressing/ spiritual outlook
	3	answer or use cell phone in the class
	4	concern his/her teaching results
	5	answer student's question during out-of-class hours
	6	polite speaking / respect students
Teaching content	7	familiar with teaching contents
	8	present the contents correctly and clearly
	9	teaching contents match study progress
	10	more than 80% in-class time for teaching
	11	directly read the contents in the textbook or PPT
	12	recommend additional materials
Teaching skills and methods	13	clearly speaking / correctly pronouncing
	14	care about the student's attendance
	15	emphasis points and summarize regularly
	16	maintain teaching order in class
	17	write on the blackboard if necessary
	18	interaction between teacher and student / encourage asking question in class
	19	homework closely related to teaching contents
	20	correct homework carefully with feedback
Teaching results	21	The feeling of how much is learned through the course
	22	how much homework can be finished by the student himself

Table -2 shows the main questions for the normal courses that the students are required to answer. There are four options for a question, i.e. very good, good, medium and poor. There are potential weights of the questions which are embedded into the proposed deep model. We design a supervised learning style strategy to train the model. Firstly we divide the dataset into three parts with sizes 1000, 2000, and 2720. We denote them as D1, D2 and D3. D1 is manually scored ranging from 0 to 100 by some experienced teachers through conversation with students to know their ground truth attitude towards the course. D2 is an unsupervised learning set for outlier validation. Normally the data records in D2 are fed into the model trained with D1 and the records having too low or too high scores are picked out for manually scoring. And we retrain the model with these outlier records with the manual scores. D3 is used to evaluate the model performance.

The experiment environment is Intel i7-920, 16GB Memory, 512GB hard-disk and a NVIDIA GTX 1080 graphical acceleration card (8GB on-board memory). The data is stored in a MySQL database and we use Matlab 2015b to implement the proposed model. The stacked auto-encoders used in the proposed model is implemented by the famous deep learning toolbox [13].

Table -3 shows the overall performance of the proposed model compared to the current statistics-based model (equally weighted each question). To get a relatively stable result, we run the algorithm 10 times and record the mean and variance of the results.

Table -3 Overall performance

	A	B
Q1	$92.4\% \pm 1.8\%$	$84.0\% \pm 2.6\%$
Q2	$91.9\% \pm 2.1\%$	$83.7\% \pm 2.9\%$
Q3	$90.5\% \pm 2.3\%$	$88.1\% \pm 3.2\%$

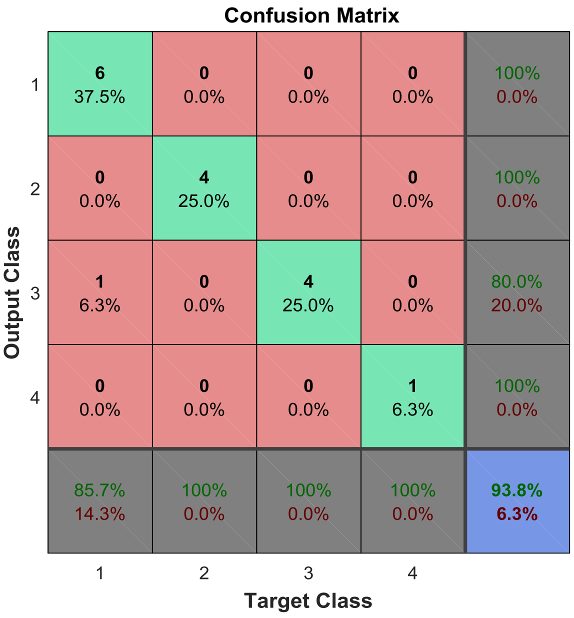
ALL	$88.6\% \pm 2.4\%$	$80.1\% \pm 3.7\%$

In Table -3, column **A** stands for the proposed method and column **B** stands for the current statistical based method. The rows **Q1**, **Q2**, **Q3** and **ALL** stand for which questionnaires are considered in evaluation. We use accuracy measurement to evaluate the model performance, which is defined as Eq. (8).

$$acc(y, y*) = 1 - \frac{|y - y^*|}{y^*} \tag{9}$$

We can see from the table that the proposed method well-performs the method for comparison and the variances are smaller which means the proposed method is more stable than the method for comparison. Both mean and variance indicate that the proposed method captures the potential principles that govern the teaching quality better and we owe this point to the power of deep models.

Meanwhile, we notice that for universities that the levels of teaching quality are meaningful, including good (90-100), fair (80-90), medium (70-80) and poor (less than 70). Hence we use a bin function to categorize the scores generated from the proposed model and report a confusion matrix to further illustrate the model performance. In such case, we gather the two questionnaires from 8 schools in our university and manually score them and compare to the output of the proposed model. Fig. 2 shows the confusion matrix in this evaluation case.



Output Class \ Target Class	1	2	3	4	
1	6 37.5%	0 0.0%	0 0.0%	0 0.0%	100% 0.0%
2	0 0.0%	4 25.0%	0 0.0%	0 0.0%	100% 0.0%
3	1 6.3%	0 0.0%	4 25.0%	0 0.0%	80.0% 20.0%
4	0 0.0%	0 0.0%	0 0.0%	1 6.3%	100% 0.0%
	85.7% 14.3%	100% 0.0%	100% 0.0%	100% 0.0%	93.8% 6.3%

Fig. 2 Confusion matrix of the proposed model for 4-class classification

In Fig. 2, the X and Y axis stand for the ground truth class and model output class labels. Confusion matrix is a powerful chart tool for illustrating the model performance of multiple classification problem. Based on the definition of confusion matrix, we can see that the overall model performance is up to 93.8%.

IV. CONCLUSION

In this paper, we proposed a deep learning based method for teaching quality assessment. We use stacked auto-encoders as the main model and design an ensemble style regressor for the output layer. The proposed model is evaluated on a dataset from our university and we get promising results compared to current statistical methods. Deep learning model is powerful in capturing potential trends and concepts which do not have explicit formulas. Its application on teaching quality assessment shows some new powerful tools for universities to accurately evaluate and improve their teaching quality.

V. REFERENCE

[1] Y. A. Feldman, "Teaching quality object-oriented programming," *J. Educ. Resour. Comput.*, vol. 5, no. 1, Mar. 2005.

[2] E. Tempero, "Experiences in teaching quality attribute scenarios," in *Proceedings of the Eleventh Australasian Conference on Computing Education - Volume 95*, ser. ACE '09. Darlinghurst, Australia, Australia: Australian Computer Society, Inc., 2009, pp. 181–188.

[3] H. Al-Mubaid, A. Abukmail, and S. Bettayeb, "Empowering deep thinking to support critical thinking in teaching and learning," in *Proceedings of the 2016 ACM SIGMIS Conference on Computers and People Research*, ser. SIGMIS-CPR '16. New York, NY, USA: ACM, 2016, pp. 69–75.

[4] C. Ardito, R. Lanzilotti, R. Polillo, L. D. Spano, and M. Zancanaro, "New perspectives to improve quality, efficacy and appeal of hci courses," in *Proceedings of the 11th Biannual Conference on Italian SIGCHI Chapter*, ser. CHItaly 2015. New York, NY, USA: ACM, 2015, pp. 188–189.

[5] P. T. Fincias, J. F. M. Izard, and P. N. Gutiérrez, "Emotional competences' development and evaluation in the non-university teaching staff in spain," in *Proceedings of the Second International Conference on Technological Ecosystems for Enhancing Multiculturality*, ser. TEEM '14. New York, NY, USA: ACM, 2014, pp. 507–512.

[6] G. Fischer and J. W. von Gudenberg, "Improving the quality of programming education by online assessment," in *Proceedings of the 4th International Symposium on Principles and Practice of Programming in Java*, ser. PPPJ '06. New York, NY, USA: ACM, 2006, pp. 208–211.

[7] Y. Wang, H. Yao, S. Zhao, and Y. Zheng, "Dimensionality reduction strategy based on auto-encoder," in *Proceedings of the 7th International Conference on Internet Multimedia Computing and Service*, ser. ICIMCS '15. New York, NY, USA: ACM, 2015, pp. 63:1–63:4.

[8] Y. Bengio, "Deep learning and cultural evolution," in *Proceedings of the Companion Publication of the 2014 Annual Conference on Genetic and Evolutionary Computation*, ser. GECCO Comp '14. New York, NY, USA: ACM, 2014, pp. 1–2.

[9] Y. Bengio, "Deep learning of representations: Looking forward," in *Proceedings of the First International Conference on Statistical Language and Speech Processing*, ser. SLSP'13. Berlin, Heidelberg: Springer-Verlag, 2013, pp. 1–37.

[10] R. Zhong and T. Tezuka, "Parametric learning of deep convolutional neural network," in *Proceedings of the 19th International Database Engineering &#38; Applications Symposium*, ser. IDEAS '15. New York, NY, USA: ACM, 2014, pp. 226–227.

[11] S. Zhou, Q. Chen, and X. Wang, "Fuzzy deep belief networks for semi-supervised sentiment classification," *Neurocomput.*, vol. 131, pp. 312–322, May 2014.

[12] Y. Li, H. Su, C. R. Qi, N. Fish, D. Cohen-Or, and L. J. Guibas, "Joint embeddings of shapes and images via cnn image purification," *ACM Trans. Graph.*, vol. 34, no. 6, pp. 234:1–234:12, Oct. 2015.

[13] R. B. Palm, "Prediction as a candidate for learning deep hierarchical models of data," Master's thesis, 2012.