ONLINE EDUCATION EVALUATION FOR SIGNAL PROCESSING COURSE THROUGH STUDENT LEARNING PATHWAYS

Kelvin H.R. Ng[⋆] *Sivanagaraja Tatinati*[⋆†] *Andy W.H. Khong*[⋆]

[⋆] Nanyang Technological University, Singapore
[†] Delta-NTU Corporate Lab, Nanyang Technological University, Singapore

ABSTRACT

Impact of online learning sequences to forecast course outcomes for an undergraduate digital signal processing (DSP) course is studied in this work. A multi-modal learning schema based on deep-learning techniques with learning sequences, psychometric measures, and personality traits as input features is developed in this work. The aim is to identify any underlying patterns in the learning sequences and subsequently forecast the learning outcomes. Experiments are conducted on the data acquired for the DSP course taught over 13 teaching weeks to underpin the forecasting efficacy of various deep-learning models. Results showed that the proposed multi-modal schema yields better forecasting performance compared to existing frequency-based methods in existing literature. It is further observed that the psychometric measures incorporated in the proposed multi-modal schema enhance the ability of distinguishing nuances in the input sequences when the forecasting task is highly dependent on human behavior.

Index Terms— Deep learning, Online education, Learning sequence, Resource usage

1. INTRODUCTION

The use of technology for education is not always affirmative. Online education such as massive open online courses (MOOC) has attracted debates surrounding its efficacy to deliver content as effective as brick-and-mortar classroom settings. It is well known that online learning systems are largely plagued with low retention rates and low knowledge transference as a result of lowered engagement and higher distraction levels [1]. Dropout rates of MOOCs in the first few weeks can be as high as 93% and the overall completion rates usually hovers only around 10% [2]. Specifically, an undergraduate MOOC DSP has observed only an average of 3% completion over multiple course runs [3]. To diagnose this problem, education data mining research focus on developing tools for the purposes of detecting undesirable behaviors and early detection of possible drop-out to intervene and prevent such destructive outcomes [4, 5].

Most of the existing tools and models in existing literature are derived from time-based features extracted from transactional logs of content access and performance records when users navigate through the course. Such information is derived at learning checkpoints or over a set period of duration or set of activities. These features have been employed to predict (a) the type of certification awarded to participants through their engagement and performances levels [6] and (b) dropout in the first three weeks using various participation measures during the first week [7]. The frequencies of interaction actions were also analyzed to understand how students consume video media and how these media can be designed to maintain student engagement [8]. While they show promising results in identifying when course-level outcomes, such as dropout and grade outcomes, did not meet desired expectations, these frequency-based predictive models are not diagnostic of learning itself and may not generalize to other learning outcomes such as knowledge mastery.

Learning itself is often considered as a process where individuals constantly make numerous spontaneous decisions based on feedbacks from internal self-monitoring mechanisms or external instructional feedbacks from educators and extrinsic motivators [9]. In an online learning platform, this manifests as a series of actions an individual takes to achieve their intended outcomes, of which may not be aligned with the intended outcome of the course. The nuances in the online transactional action logs indicative of individuals motivation and goal to engage in such an online system are washed out when they are aggregated as frequency-based features during preprocessing. The study of learning action sequences can take advantage of sequence-based predictive models that learn from these nuances to perform the predictions. In this aspect, Fei et al. trained a recurrent neural network (RNN) with the frequencies of various teaching modes that span across a week to predict dropout [10]. These teaching modes include lectures, forums, and quizzes. However, learning education sequences are beyond simply adapting existing deep learning techniques. Tang et al. were unable to achieve good prediction results when they trained a long short-term memory (LSTM) network using raw sequences [11].

Deep architectures have the ability to extract appropriate features from raw inputs to perform regression and classification. The deep architecture enables the network to learn higher-level representations from lower-level features of the input data so as to identify the underlying relation between input and output spaces. RNN can be considered as having a deep architecture since the hidden states are recursively used to embed the input sequence. RNNs have seen many successful applications in text-mining and web usage mining recently. Some of these models include the classification of movie reviews for sentiment analysis [12], image captioning [13] and classification of text documents genres. Although texts and online learning sequences have fundamentally similar constructs, their compositions are different. Text documents consist of words and phrases sparsely sampled from a very large vocabulary and the semantics varies depending on how they are used. On the other hand, learning sequences are formed by highly repetitive discrete symbols from a smaller bank of possible actions and there exists no semantics of a certain action when studied in isolation. The dense vocabulary meant that the relationship between timesteps may not be independent and models with such assumption may fail.

This work was conducted within the Delta-NTU Corporate Lab for Cyber-Physical Systems with funding support from Delta Electronics Inc and the National Research Foundation (NRF) Singapore under the Corp Lab@University Scheme.

Apart from online learning sequences, a learners personality can drive his efforts towards a goal. In earlier works for learning analytics, studies were carried out to understand the impact of self-reported psychometric attributes such as motivation and persistence on online learning and how these attribute-behavior relationships changed from traditional classroom context to the cyber world. The use of technology also enables researchers to gain more insights into learning behaviour beyond static traits. Detectors of dynamic behavior such as gaming the system and off-task behaviour were developed and studied for their impact on learning outcomes [14, 15].

In this paper, we demonstrate the advantage of using sequences to train models compared to frequency-based models for the prediction of grade performances in a digital signal processing (DSP) course. The models will be trained using clickstream logs of lecture content viewing when learners prepare for their weekly tutorials. More importantly, we show that the inclusion of static psychometric attributes of students improved model prediction accuracy.

2. THE PROPOSED LSTM-BASED PREDICTION MODEL

The proposed architecture consists of four main components: the input layer, convolution layer, LSTM, and psychometric covariates.

Input layer: Without any pre-processing, neural networks cannot accept discrete learning action symbols as its inputs. One of the methods to convert these discrete symbols into a vector is to use one-hot encoding where a numerical Boolean indicator identifies the word in a n-length vector describing all possible symbols in the models. The sparsity and lack of information makes it difficult for a neural network to learn features that are required to perform the prediction task. Two popular approaches to overcome such sparsity involve the use of either an auto-encoder or an embedding layer to convert the sparse vector into a dense continuous vector.

A typical auto-encoding input layer consists of a fully-connected feed-forward neural network that compresses the inputs to much smaller dimensions across a few hidden layers and subsequently using the hidden representation to reconstruct its input. The hidden states then form the dense continuous vector representation of the discrete input. Auto-encoders are typically trained prior to the modelling and continue to be updated together with other components of the model.

On the other hand, an embedding layer consists of an $n \times m$ lookup table with randomly initialized values. Each of the n rows in the lookup table is a m-dimensional vector representing an input. The values of the lookup table are updated during training using information propagated back from the model objective function.

In this paper, we compare the impact of different learning paradigms of the embedding and auto-encoding layer towards modeling education sequences.

Convolution layer: A convolution layer consists of a set of filters (kernels) that is much smaller than the matrix that it is applied onto. Each kernel learns local spatial information in a matrix by repeatedly convoluting across the matrix at every k strides. The convolution layer can be used to extract and aggregate patterns of a dense matrix into a more concise representation which a neural network can better utilize to perform the prediction.

LSTM: The LSTM recurrent neural network is a variant of the RNN family where the hidden states of the neural network are reused recursively to learn the appropriate time-series features of a sequence pattern. It is well known that RNNs suffer from training difficulty due to exploding/vanishing gradients where gradients propagated through the network either amplifies or attenuates with each iteration, resulting in instability since large changes are made to the hidden states. LSTMs address this issue by using three control gates (input, output, forget gates), to create a gradient-neutral hidden state for the network where the determinant of its gradients for the hidden cell is bounded.

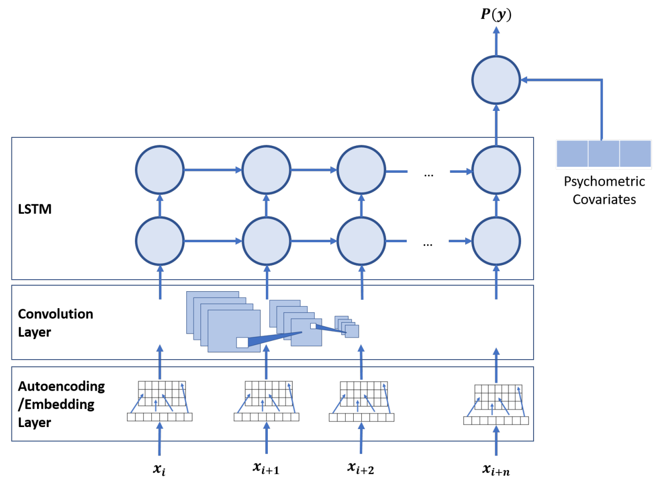

Fig. 1. Proposed architecture.

Covariate layer: To include the psychometric measures, the measures are z-normalized and fed to a fully-connected layer, of which the outputs of the layer are concatenated to the last hidden state of the LSTM cell. The concatenated vector then forms the hidden representation of the conditional probability of the prediction task where a final fully-connected layer performs the prediction. The architecture is illustrated in Fig. 1.

3. DATA COLLECTION

Data is collected from an undergraduate level DSP course covering topics such as linear time-invariant systems, Fourier transforms and basic filter properties and designs. The course is taught over 13 teaching weeks using flipped-classroom pedagogy. Each week students are required to prepare for the class through a set of online learning materials and thereafter expected to participate in a tutorial where a quiz will be administered at the start of the class to assess students readiness levels. Instructors then facilitate a discussion to clarify and reinforce relevant topics. The learning materials are hosted through the online LAMS learning management system where a linear sequence for each set of content content can be administered weekly. These materials consist of video lectures, interactive media contents, multiple-choice practice questions, open-ended response questions as well as forum discussions. The practice questions are optional and students are prompted to switch over from the content sequence to perform the practice questions when they reached a checkpoint. They may proceed with the learning materials first without performing the practice tasks. These practices only serve as a self-monitoring tool and students are conveyed that this will not affect their course outcome. In the 4th and 6th weeks of the course, students have an optional hands-on class for MATLAB in place of the tutorial and therefore do not have the in-class quiz. On the 7th week, there is mid-term assessment to assess the content that has been covered thus far. This quiz will contribute to the students course outcome.

The data is collected over the first six weeks of the course in the form of transaction logs where the system records how students interact with the video lecture resources and practice questions to prepare for their weekly tutorials. At the beginning of the course, we also requested students to perform a 44-items questionnaire evaluating 6 personality traits. The traits have been validated with the study population. The list of clickstream actions and psychometric attributes can be found in Table 1. A total of 148 students participated in this study.

Content Type	Type of information collected
Video	Start, Pause, Resume, End video, End activity, Fast-forward, Enter Fullscreen, Exit Fullscreen, Speed up, Speed down
Multiple-choice Question	Select choice, Edit choice, Submit answers
Personality Traits	Grit [16], Theory of intelligence [17], Academic Bouyancy [18], Performance-approach goals, Performance-avoidance goals, Mastery goals [19]

Table 1. List of input features.

3.1. Validity of performance scores

A visual analysis of student performance on the online LAMS practice questions in Fig. 2 shows that while there are variations in scores during earlier practice questions, majority of the students are answering the online practices perfectly after the 8th checkpoint. This is consistent with commonly observed online settings where questions are delivered immediately after the relevant content. These multiple-choice questions (MCQ) are often intended to serve as a monitoring mechanism rather than an assessment of ones knowledge and therefore are of relatively low complexity and specificity. Metrics derived from these assessments do not correlate well with individual knowledge mastery and also do not have sufficient statistical variation to make predictive task meaningful. On the other hand, questions posed in the weekly in-class quiz before the tutorial consist mainly of application questions and students are required to pen down the steps involved within a limited time frame. These weekly in-class quizzes posed sufficient complexity and difficulty to generate varied performance results as illustrated in Fig. 3 and are therefore chosen as our prediction metric to evaluate student knowledge levels. Information on how students prepare for the tutorial is explicitly captured in the data for the first 4 weekly in-class quizzes and the mid-term assessment. For students who missed their tutorials, they do not have their weekly quiz scores for the week and these missing values are assigned the cohort average for that week. The average is chosen in lieu of zeroing the scores to minimize the impact on the distribution of scores.

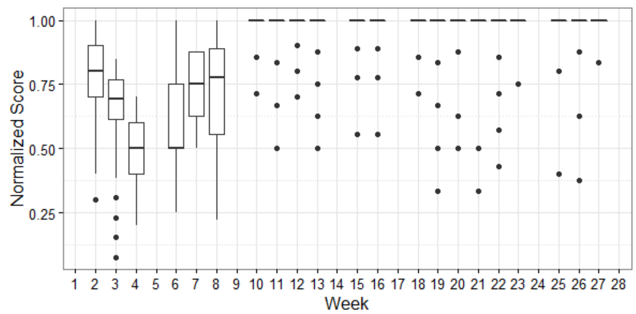

Fig. 2. Distribution of scores for online practice questions.

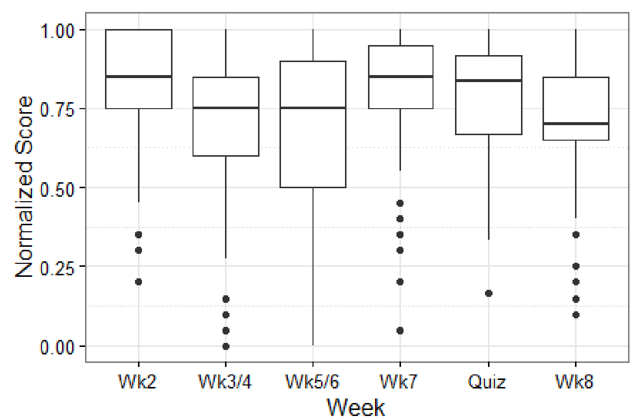

Fig. 3. Distribution of scores for in-class readiness assessment.

4. EXPERIMENTS

In this comparison analysis, two variants of sequence-based methods and two variants of frequency-based models are employed for the data when all past weekly preparations are considered. Under the category of frequency-based models, 1) a multi-layer perceptron (MLP) with 3 hidden layers each with 100 hidden nodes and tanh activation functions and 2) convolutional neural network (CNN) with three layers of 500, 300, and 100 feature maps with 3×3 filters convolution layers are considered. For compactness, the MLP method is denoted as F1 and CNN method as F2. As the sequence-based models, we employed four LSTM based network structures for forecasting the students performance. The LSTM models, denoted as S1, S2, S3, S4, have 100 hidden dimensions each. For S1 and S3, the input layer consists of a 50-dimension auto-encoding layer while S2 and S4 each has a 50-dimension embedding layer for its input layer. For S3 and S4, a convolution layer of thirty-two 7×7 filters are applied on the input layer before passing the outputs to the LSTM cells. All the models are implemented in Tensorflow and are trained using a 60/20/20% data partition for training, validation, and testing sets, respectively. All the models are trained for 350 epochs or stopped when the validation losses start to diverge from training losses. The performances of the models are evaluated using the root-mean-squared errors of the weekly score prediction errors of the testing dataset.

4.1. Impact of feature extraction for sequence input

In this work, auto-encoding and embedding layers are chosen to extract features from the raw input sequences in the sequence-based models. Fundamentally these two processes learn through different mechanisms. The auto-encoding layer learns to reconstruct its input by reducing the entropy of the auto-encoder for the required task instead of random initialization, and subsequently trains the learnt feature set for the forecasting task at hand. The embedding layer, on the other hand, learns through complete random permutation and

Data Type	Model	Training Loss (RMSE)	Validation Loss (RMSE)	Testing Loss (RMSE)
Single-week frequency-based	F1	0.269/ 0.223	0.250/ 0.219	0.250/ 0.305
Single-week sequence-based	S1	0.129/ 0.117	0.408/ 0.453	0.331/ 0.232
	S2	0.158/ 0.179	0.569/ 0.505	0.342/ 0.235
	S3	0.0870/ 0.113	0.401/ 0.432	0.331/ 0.224
	S4	0.212/ 0.136	0.361/ 0.434	0.285/ 0.237
Full-history frequency-based	F1	0.968/ 1.071	1.052/ 1.181	1.069/ 1.248
	F2	0.0345/ 0.0272	0.382/ 0.328	0.323/ 0.243
Full-history sequence-based	S1	0.0345/ 0.0272	0.382/ 0.328	0.323/ 0.243
	S2	0.0347/ 0.0345	0.299/ 0.328	0.311/ 0.227
	S3	0.0117/ 0.0404	0.276/ 0.300	0.304/ 0.223
	S4	0.0281/ 0.0291	0.312/ 0.312	0.307/ 0.251

Table 2. Score predictions without/with the use of psychometric covariates.

updates from gradient descent based on an objective function error rate. The two methods are expected to extract slightly different features, but they do not have significant impact on the models performances. The training errors for auto-encoding layer models (S2 and S4) are lower than embedding layer models (S1 and S3) but generalize better with unseen data compared to the latter. For completeness, we report that embedding layer models achieved testing RMSE as small as 0.331 and 0.304 for single-week based and full-history based sequence forecasting models, respectively. We observed similar trend regardless of whether psychometric covariates are considered in the input feature. However, models with psychometric measures achieved better performance compared to the models without those measures, as depicted in Table 2. From these results, we infer that although the online learning sequences do not have a single distribution all the time, the proposed architecture with covariates attempts to learn the underlying patterns effectively and yielded better forecasting performance.

4.2. Impact of observation period on forecasting performance

Frequency-based model performances are adversely affected when the period of consideration is extended. One of the possible reasons could be due to the input feature set distribution collapsing into a unitary mode. Specifically, when the preparation history is aggregated, the model no longer differentiates students with consistent efforts from those who sporadically spend much time on online materials to catch up with the curriculum. The only information left that the model could utilize is to predict performance scores based on whether students prepared for class. Most of the sequence-based models, however, show improvement in the prediction error. For single-week based experiments, RMSE of 0.285 and 0.250 are obtained with sequence-based models and frequency-based models, respectively, whereas in full-history based methods the improvement is significant - the RMSE is 0.304 with sequence-based methods and 0.328 with frequency-based methods, as shown in Table 2. The use of sequences rather than frequencies meant that continuous evaluation is possible at any instance rather than checkpoints along the curriculum. With the obtained forecasting performance, educators can provide just-in-time help to students that requires their attention while challenging students who are performing well to go beyond the curriculum in order to achieve higher metacognitive skills by generalizing knowledge obtained into other applicable domains. With sufficient modes of delivery, personalized learning can also be achieved by evaluating how students react to different types of media. Students may not always follow strictly to the curriculum timeline. When models are provided with the raw information during training, they can identify some of these irregularities as non-disruptive behaviours towards the individuals performance. The impact of different sequence nuances on the performance is outside the scope of the paper.

4.3. Impact of psychometric covariates for sequence-based models

Sequences can exhibit many varieties, some with the same given outcome. We hypothesized earlier that some of the psychometric measures could serve as covariates to assist models in identifying important nuances in the sequences leading to better prediction performances. Indeed, most of the models other than F1 performed better with the inclusion of the psychometric measures, as shown in Table 2. The covariates are chosen to cover a broad spectrum of students describing how they are driven by their motivation and goals. Results further highlighted that the existence of covariates helped in reducing the input feature redundancy and thus any method with covariates yielded less RMSE compared to the models without covariates. It can be seen from our results that the sequence-based models, irrespective of auto-encoding or embedding, utilize the information obtained with psychometric measures achieve higher prediction performances compared to the frequency-based methods.

5. CONCLUSION

We analyzed the forecasting performance of frequency-based models to evaluate learning efficacy of a digital signal processing course in an online environment. We also demonstrated that aggregation of features resulted in a loss of information that could be diagnostic of learning. We hypothesized that nuances in the learning pathways are important in identifying learning outcomes of individuals. In this aspect, we demonstrated that models learnt through the complete sequences tend to perform better than frequency-based models. Furthermore, we have shown that psychometric covariates contain information crucial for the identification of semantics of the nuances in the learning sequences. While the use of sequence-based models and psychometric covariates show promising results, the accuracy of these models have the potential to improve the learning and teaching procedures for DSP courses.

6. REFERENCES

[1] T. Y. Yang, C. G. Brinton, C. Joe-Wong, and M. Chiang, "Behavior-based grade prediction for MOOCs via time series neural networks," *IEEE Journal of Selected Topics in Signal Processing*, vol. 11, no. 5, pp. 716–728, 2017.

[2] D. F. O. Onah, J. Sinclair, and R. Boyatt, "Dropout rates of massive open online courses: behavioural patterns," in *Proc. Int'l Conf. Education and New Learning Technologies*, 2014, pp. 5825–5834.

[3] T. A. Baran, R. G. Baraniuk, A. V. Oppenheim, V. Alan, P. Prandoni, and M. Vetterli, "MOOC adventures in signal processing: Bringing DSP to the era of massive open online courses," *IEEE Signal Processing Magazine*, vol. 33, no. 4, pp. 62–83, 2016.

[4] D. Yang, T. Sinha, D. Adamson, and C.P. Rosé, "Turn on, tune in, drop out: Anticipating student dropouts in massive open online courses," in *Proc. NIPS Workshop Data Driven Education*, 2013, vol. 11, p. 14.

[5] S. Halawa, D. Greene, and J. Mitchell, "Dropout prediction in MOOCs using learner activity features," in *Proc. 2nd European MOOC Stakeholder Summit*, 01 2014, vol. 37, pp. 7–16.

[6] S. Jiang, A. Williams, K. Schenke, M. Warschauer, and D. O'dowd, "Predicting MOOC performance with week 1 behavior," in *Proc. Int'l Conf. Educational Data Mining*, 2014.

[7] C. Ye, J. S. Kinnebrew, G. Biswas, B. J. Evans, D. H. Fisher, G. Narasimham, and K. A. Brady, "Behavior prediction in MOOCs using higher granularity temporal information," in *Proc. 2nd ACM Conf. Learning @ Scale*. ACM, 2015, pp. 335–338.

[8] N. Li, L. Kidzinski, P. Jermann, and P. Dillenbourg, "MOOC video interaction patterns: What do they tell us?," 2015.

[9] G. Krudysz and J. McClellan, "Study-flow: Studying effective student-content interaction in signal processing education," in *Proc. IEEE Int'l Conf. Acoustics, Speech, Signal Proc.*, 2017, pp. 6354–6358.

[10] M. Fei and D. Y. Yeung, "Temporal models for predicting student dropout in massive open online courses," in *Proc. IEEE Int'l Conf. Data Mining Workshop*, 2015, pp. 256–263.

[11] S. Tang, J. C. Peterson, and Z. A. Pardos, "Deep neural networks and how they apply to sequential education data," in *Proc. 3rd ACM Conf. on Learning @ Scale*. ACM, 2016, pp. 321–324.

[12] M. Wöllmer, F. Weninger, T. Knaup, B. Schuller, C. Sun, K. Sagae, and L. P. Morency, "Youtube movie reviews: Sentiment analysis in an audio-visual context," *IEEE Intelligent Systems*, vol. 28, no. 3, pp. 46–53, 2013.

[13] A. Karpathy and F. F. Li, "Deep visual-semantic alignments for generating image descriptions," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 2015, pp. 3128–3137.

[14] R. S. J. Baker, "Modeling and understanding students' off-task behavior in intelligent tutoring systems," in *Proc. SIGCHI Conf. Human Factors in Computing Systems*. ACM, 2007, pp. 1059–1068.

[15] J. A. Walonoski and N. T. Heffernan, "Detection and analysis of off-task gaming behavior in intelligent tutoring systems," in *Proc. Int'l Conf. Intelligent Tutoring Systems*. Springer, 2006, pp. 382–391.

[16] A. L. Duckworth, C. Peterson, M. D. Matthews, and D. R. Kelly, "Grit: perseverance and passion for long-term goals," *Journal of personality and social psychology*, vol. 92, no. 6, pp. 1087–101, 2007.

[17] C. S. Dweck, *Self-theories: Their role in motivation, personality, and development*, Psychology Press, 2000.

[18] A. J. Martin and H. W. Marsh, "Academic resilience and its psychological and educational correlates: A construct validity approach," *Psychology in the Schools*, vol. 43, no. 3, pp. 267–281, 2006.

[19] A. J. Elliot and M. A. Church, "A hierarchical model of approach and avoidance achievement motivation," *Journal of personality and social psychology*, vol. 72, no. 1, pp. 218, 1997.