



The Potential for the Use of Deep Neural Networks in e-Learning Student Evaluation with New Data Augmentation Method

Andrzej Cader^(✉)

Information Technology Institute, University of Social Sciences, Lodz, Poland
acader@san.edu.pl

Abstract. This study attempts to use a deep neural network to assess the acquisition of knowledge and skills by students. This module is intended to shape a personalized learning path through the e-learning system. Assessing student progress at each stage of learning in an individualized process is extremely tedious and arduous. The only solution is to automate assessment using Deep Learning methods. The obstacle is the relatively small amount of data, in the form of available assessments, which is needed to train the neural network. The specificity of each subject/course taught requires the preparation of a separate neural network. The paper proposes a new method of data augmentation, Asynchronous Data Augmentation through Pre-Categorization (ADAPC), which solves this problem. It has been shown that it is possible to train a very effective deep neural network with the proposed method even for a small amount of data.

Keywords: Deep Learning · e-learning · Deep neural networks · New data augmentation method

1 Introduction

Deep Learning (DL) methods in teaching began to spread after 2010 [1–3]. In recent years, a significant increase in the use of neural networks in teaching has been seen [4–6], and also in the field of student evaluation automation [7–9]. Two areas that automation applies to can be distinguished. The first relates to automated essay scoring and the second to automatic short answer grading, automatically classifying student responses as correct or not, based on a set of previous correct answers [10, 11]. Particularly interesting are attempts to use DL capabilities in the field of text analysis [12, 13]. Methods based on the use of recurrent neural networks [14–16], including bidirectional LSTM networks [5], dominate here.

The priority of modern education is to adapt the methods and pace of knowledge and skills transfer to the individual predispositions of each individual student. Such a strategy requires both the division of the entire learning process into small multi-variant stages, and also the assessment of the level of mastery of knowledge and skills at the end of each stage. It is possible to shape the course of the entire teaching process for

each student separately by using assessment that is carried out in stages. Such multi-variability of choice of further educational path is important – the choice of the type of next stage from among several options available, based on the result of the previous stage’s evaluation. The assessment of a particular stage should be derived from many assessments that occur during various activities. These grades should be grouped under specific validation areas, e.g. test grades, practical tasks, own work, project grades, etc. The source of grades can be teachers or other students as part of group work, or it can be a self-assessment. Assessments can also come from automatic validation systems – automatic test evaluation, automatic text, image, speech, etc. The validation process in this system concept is very tedious and extremely burdensome for the tutor leading a given group of students – many rated persons, a very large number of stages, often very limited contact with the assessed student, many grades from various sources. In such a situation it is difficult to decide what final grade to make. It seems that in such a situation it is optimal to use an automatic system based on a properly trained neural network.

2 Comprehension and Data Preparation

This work presents the research stage of a broader program related to the development of a platform for personalized education of students at the University. Its purpose is to explore the possibility of creating a system for automatic validation of the teaching stages of a selected subject using DL methods. It is assumed that a deep neural network will be trained based on a small set of training data - student assessments.

The designed neural network should take into account the context defined by the environment in which the evaluation will take place. The specificity of assessment depends primarily on the structure and content transmitted in the educational process and the type of competences acquired by the student. In other words it depends on the subject being taught. Moreover, this condition will be determined by the specific curriculum, the assumed teaching objectives and even by different ways of organizing classes and the profile of the teaching staff. This means that in each specific case, training the neural network should be adapted to the conditions presented above. This leads to a significant reduction in the amount of training data available. In this case, it is difficult to use existing methods of data augmentation [17–19]. One of the possibilities is to use the properties of the student grade set, which was referred as data asynchronism.

Def. Asynchronous data - a set of data whose ranking (order) does not affect the information contained in this set. In particular, asynchronous data does not form a time series or sequence ordered in a different way in time or space.

From this definition, it follows that the set of feature values (grades) that determine the state of the student’s knowledge and skills is a set of asynchronous data. The grades determine level of students mastery, to a large extent, regardless of the order in which they occur. Of course, this is some simplification resulting from the assumed model.

Lemma. Let B be a discrete set of N features describing the state of a given object: $B = \{c_1, c_2, \dots, c_N\}$ and which can take a finite number of v_{ij} values (i - feature number, j - number value). If all v_i sets are asynchronous data sets, then each combination of individual elements selected from each v_i set reflects a certain state of the object.

It follows from the above that for asynchronous data relating to object feature values, each combination of individual feature values can be an input vector of the neural network classifying the object's state. It should be clarified that individual combinations correspond to the detailed states, while the sequences of values of the attributes v_i represent the generalized state. Thus, by presenting many detailed vectors of the neural network, we are building a representation of the generalized state. The number of input vectors for each dataset is the product of the number of elements in each feature v_i .

A group of 80 students was selected for the experiment, whose grades generated training data for the neural network and a separate group of 40 students for the test set. Assessments were collected as part of the subject of physics in computer science at the University of Social Sciences in Lodz. Scores on a scale of 1 to 10 (0 means no rating) were issued in 12 categories: 1. Ability to create written studies; 2. Ability to prepare projects; 3. Level of solving theoretical sentences; 4. Ability to solve practical problems; 5. Ability to solve tests; 6. Substantive formulation of the oral answer; 7. Participation in the discussion and substantive activity; 8. Participation in consultations; 9. Own work; 10. Creativity; 11. Cooperation as part of group tasks; 12. Timeliness of tasks. The output of the trained network (labels) were the final grades issued by the tutor at the end of the semester (Table 1).

Table 1. Example of student assessments used to train the network.

Id	Labels	Cat_1	Cat_2	Cat_3	Cat_4	Cat_5	Cat_6	Cat_7	Cat_8	Cat_9	Cat_10	Cat_11	Cat_12
1	9	9, 8, 8	10, 8, 9	9, 9, 9	8, 8, 10	9, 8	9	7	5, 7	6, 8	9	9, 8	9, 9, 8
2	8	8, 9, 9	9, 6, 10, 7	8, 5, 10	7, 8, 8	9, 10	9	3	5, 8, 9	8, 7	0	6, 7	9, 10, 9
3	3	1, 2, 2, 4	3, 4, 2	2, 3	5, 4, 6	1, 9	2	1	1, 2	2, 1	2	1	5, 4, 3
4	4	7, 5, 5	5, 3	5, 4, 3, 6	7, 6	4, 6	4	4, 2	2, 4	3, 4	4	5	7, 8

Preparation of training data (80 students) included the following stages:

1. Assembling of all combinations of grades from Cat_1 to Cat_12 (one grade from each field) with the assignment of each combination of the same label, separately for each student (Id)
2. Random shuffle of all combinations
3. Separation of the set into train_data and train_labels and standard preparation of input data with normalization train_data.

560 688 training data were obtained using the procedure presented. At the stage of selecting the network model and tuning, a set of 160,000 validation data was temporarily separated from the training data. Test data were prepared on the basis of assessments of a separate group of 40 students. Test vectors were built from an average of individual categories rounded to the total value.

3 Experimental Results

Various models of neural networks and hyperparameter sets were considered in the validation process. The optimal turned out to be the use of a fully connected neural network with five dense layers. In layers 1 to 5, the *ReLU* activation function was used, and *Softmax* used in the output layer. The output layer neurons correspond to trained categories, which are final grades, expressed on a point scale from 0 to 10. The total number of parameters (weights and biases) was 84,043, all trained. The errors were computed based on *categorical cross-entropy* loss function and the *Adam* optimizer. Optimal mini-batch size = 100 selected. During NN training, it was determined that there was no need for regularization techniques. It is true that after 14 epochs, the effect of overfitting appeared, but up to this point the model obtained a surprisingly high training accuracy of 0.9982 (Fig. 1).

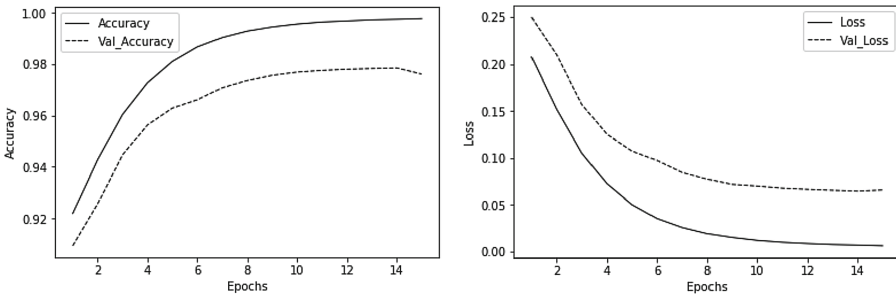


Fig. 1. Accuracy and loss function values calculated on the training (Accuracy, Loss) and validation (Val_Accuracy, Val_Loss) sets.

During testing, the results of prediction of the trained NN model were compared with the assessments proposed by the tutors. Because the *Softmax* output layer creates a probability distribution for individual categories (grades), the winning category is the one with the highest probability value. Out of 40 evaluated in 33 cases, the predictors were fully compatible with tutors' assessments. In four cases, the value of the prediction differed by one point from the tutor's assessment, in two by 2 points and in one by 4 points.

4 Conclusion

It has been shown that it is possible to use a deep neural network for extremely small amounts of data if they meet the asynchronous condition, i.e. independence of the way they are ordered. In this case, you can use a new method of data augmentation, which is technically called Asynchronous Data Augmentation through Pre-Categorization (ADAPC). Based on this method, you can train a medium-sized neural network that effectively classifies student achievement in the relatively narrow area of one subject

(course) or module. This creates the possibility of quick and easy generation of artificial structures for automatic validation of educational processes. It should be emphasized that the ADAPC method can be used in many other areas in both classification and regression issues, provided that the processed data has the asynchronous feature. The model has been developed to meet the needs of a larger e-learning system as a link in profiling the individual education path of university students.

References

1. Romero, C., Ventura, S.: Educational data mining: a review of the state of the art. *IEEE Trans. Syst. Man Cybern. Part C Appl. Rev.* **40**(6), 601–618 (2010)
2. Romero, C., Ventura, S.: Data mining in education. *Wiley Interdiscipl. Rev. Data Mining Knowl. Discov.* **3**(1), 12–27 (2013)
3. Pena-Ayala, A.: Educational data mining: a survey and a data mining-based analysis of recent works. *Expert Syst. Appl.* **41**(4), 1432–1462 (2014)
4. Hu, B.: Teaching quality evaluation research based on neural network for university physical education. In: 2017 International Conference on Smart Grid and Electrical Automation (ICSGEA), pp 290–293. IEEE (2017). <https://doi.org/10.1109/ICSGEA.2017.155>
5. Kose, U., Arslan, A.: Optimization of self-learning in computer science engineering course: an intelligent software system supported by artificial neural network and vortex optimization algorithm. *Comput. Appl. Eng. Educ.* **25**, 142–156 (2017)
6. Lau, E.T., Sun, L., Yang, Q.: Modelling, prediction and classification of student academic performance using artificial neural networks. *SN Appl. Sci.* **1**, 982 (2019). <https://doi.org/10.1007/s42452-019-0884->
7. Zhao, S., Zhang, Y., Xiong, X., Botelho, A., Heffernan, N.: A memory-augmented neural model for automated grading. In: Proceedings of the Fourth ACM Conference on Learning @ Scale, pp. 189–192. ACM, Cambridge (2017)
8. Alvarado, J.G., Ghavidel, H.A., Zouaq, A., Jovanovic, J., McDonald, J.: A comparison of features for the automatic labeling of student answers to open-ended questions. In: Proceedings of the 11th International Conference on Educational Data Mining (2018)
9. Sales A., Botelho A., Patikorn T., Heffernan N.T.: Using big data to sharpen design-based inference in A/B tests. In: Proceedings of the 11th International Conference on Educational Data Mining (2018)
10. Taghipour, K., Ng, H.T.: A neural approach to automated essay scoring. In: Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, pp. 1882–1891. Association for Computational Linguistics, Austin (2016)
11. Zhang, Y., Shah, R., Chi, M.: Deep learning + student modeling + clustering: a recipe for effective automatic short answer grading. In: Proceedings of the 9th International Conference on Educational Data Mining, pp. 562–567 (2016)
12. Choi, H., Wang, Z., Brooks, C., Collins-Thompson, K., Reed, B.G., Fitch, D.: Social work in the classroom? A tool to evaluate topical relevance in student writing. In: Proceedings of the 10th International Conference on Educational Data Mining (2017)
13. Neethu, G., Sijimol, P.J., Surekha, M.V.: Grading descriptive answer scripts using deep learning. *Int. J. Innov. Technol. Explor. Eng.* **8**(5), 991–996 (2019)
14. Tang, S., Peterson, J.C., Pardos, Z.A.: Deep neural networks and how they apply to sequential education data. In: Proceedings of the Third ACM Conference on Learning @ Scale (L@S 2016), pp. 321–324. ACM, New York (2016)

15. Okubo, F., Yamashita, T., Shimada, A., Ogata, H.: A neural network approach for students' performance prediction. In: Proceedings of the Seventh International Learning Analytics & Knowledge Conference (LAK 2017), pp. 598–599. ACM, New York (2017)
16. Wang, L., Sy, A., Liu, L., Piech, C.: Learning to represent student knowledge on programming exercises using deep learning. In: Proceedings of the 10th International Conference on Educational Data Mining, Cambridge, MA, USA, pp. 201–204 (2017)
17. Aggarwal, C.C.: Recommender Systems. Springer, Cham (2016)
18. Shorten, C., Khoshgoftaar, T.M.: A survey on image data augmentation for deep learning. *J. Big Data* **6**, 60 (2019). <https://doi.org/10.1186/s40537-019-0197-0>
19. Tan, Y.K., Xu, X., Liu, Y.: Improved recurrent neural networks for session-based recommendations. In: Proceedings of the 1st Workshop on Deep Learning for Recommender Systems, pp. 17–22 (2016)