Study, Attend and Predict: Academic Performance Prediction using Transformers

Nicolas Araque Volk
Facultad de Ingenieria
Universidad Metropolitana
Caracas, Venezuela
naraque@unimet.edu.ve

Mario Catapano
Facultad de Ingenieria
Universidad Metropolitana
Caracas, Venezuela
mario.catapano@correo.unimet.edu.ve

Massimo Di Gennaro
Facultad de Ingenieria
Universidad Metropolitana
Caracas, Venezuela
digennaro.massimo@correo.unimet.edu.ve

1   Abstract Proposal

Education in Latin America is a key factor for social development and poverty, but there are several barriers that stand in the way of a better education system [1]. Build tools that empower students to improve their performance in the educational system is an important task to provide a better future for all of us.

The educational data mining discipline has developed tools and techniques for educational institutions to improve their services. Predicting academic performance and dropout students enables institutions to help students that may have not graduated without help recover from their situation [8].

Predicting academic performance for on-site university settings has been also studied with recommender system [5], machine learning [9], and deep learning techniques [6][7]. The development of tools that can help students understand their performance and how to improve it is a less investigated topic.

The deep learning discipline has new developments that can help to improve the accuracy and reach of educational data mining techniques that help Latin American students to improve their educational path. Specifically, this abstract focuses on the transformer architecture for these developments.

The transformer architecture has come to change the way how natural language processing is understood from the recurrent neural networks standpoint, making it possible to train models on time series data and obtaining state of the art results. The use of this architecture has also been applied outside the NLP domain for a time series forecasting tool [2].

The objective of this research is to create a model that uses the transformer architecture for the academic performance prediction task to obtain state of the art results and use it as a recommender system for university students. We're aiming to outperform previous results where LSTM was used to predict academic performance [7]. This recommender system is going to help students decide how many and which courses to enroll in the net term. This support is going to improve the decision-making process of university students to take better and lower risk decisions.

As we focus on Latin America education, we're developing our research in the Universidad Metropolitana from Caracas, Venezuela. The first stage of the research centered on the System Engineering mayor and are the preliminary results that are stated in this abstract.

Table 1: Confusion matrix

	Passed Grade	Failed Grade
Passed Grade Predicted	1359	423
Failed Grade Predicted	204	390

We collected a dataset containing 822 different students over a 7 year period from 2011 to 2017. This dataset represents the academic performance history, containing the courses and grades obtained by these students in each term consolidating 6053. We split this dataset between train and validation data, with 739 students for training and 83 used for validation. As the split was done at a student level, there is no course data from any student on both train and validation.

We then design the prediction task as follow: given the grades of all courses enrolled in every past term, we want the model to predict the probability of obtaining a passing grade on each course the students want to enroll in the next term. This output results in useful information regardless what courses had a low probability of been passed by the student seen in the context of the overall difficulty of the term.

In order to use the academic performance history on the transformer architecture, we preprocess the data to be a dense input vector representation of all the courses concatenated with the respective grades. We then use the encoder section to generate a binary prediction for every course on the courses selected for the next term [8].

No pre-trained course embeddings were available for the use of the transformer architecture, our solution was to train random pre-defined embeddings as part of the overall training process.

The preliminary results for this research are promising. We have achieved an accuracy of 73.6 % and an F1 score of 81.26 % for the validation set. The confusion matrix is presented in table 1. The class imbalance makes the pass class have better performance than the failed class.

Several improvements can be made to this model and is a work-in-progress. We want to improve the prediction for the failed class, this would give better information to students to make a decision on low probabilities prediction courses. Also, we want to incorporate new classes to our output, making it possible to students to not only know if they're going to pass or failed a course but have a better understanding of the predicted grade.

From this result, we can conclude that the transformer architecture can be used to predict academic performance on a university dataset. Future works need to be done in order to compare with a benchmark to state the improvement with already existing techniques.

References

[1] "Education and Poverty in Latin America," Harvard.edu, 2020. https://revista.drclas.harvard.edu/book/education-and-poverty-latin-america (accessed Oct. 18, 2020). [2] Wu, N., Green, B., Ben, X., Oapos;banion, S. (n.d.). Deep Transformer Models for Time Series Forecasting: The Influenza Prevalence Case. Retrieved from https://arxiv.org/pdf/2001.08317.pdf

[3] A. Hernández-Blanco, B. Herrera-Flores, D. Tomás, and B. Navarro-Colorado, "A Systematic Review of Deep Learning Approaches to Educational Data Mining," Complexity, vol. 2019, pp. 1–22, May 2019, doi: 10.1155/2019/1306039. [4] M. Sweeney, H. Rangwala, J. Lester, and A. Johri, "Next-Term Student Performance Prediction: A Recommender Systems Approach," 2016. [Online]. Available: https://arxiv.org/pdf/1604.01840.pdf.

[5] N. Araque, G. Rojas, and M. Vitali, "UniNet: Next Term Course Recommendation using Deep Learning," arXiv.org, 2020. https://arxiv.org/abs/2009.09326 (accessed Oct. 16, 2020).

[6] N. Mduma, K. Kalegele, and D. Machuve, "A Survey of Machine Learning Approaches and Techniques for Student Dropout Prediction," Data Science Journal, vol. 18, 2019, doi: 10.5334/dsj-2019-014.

[7] Z. Iqbal, J. Qadir, Mian, Adnan Noor, and F. Kamiran, "Machine Learning Based Student Grade Prediction: A Case Study," arXiv.org, 2017. https://arxiv.org/abs/1708.08744 (accessed Oct. 18, 2020).

[8] A. Vaswani et al., "Attention Is All You Need," arXiv.org, 2017. https://arxiv.org/abs/1706.03762 (accessed Oct. 18, 2020).