Natural language processing for understanding classroom dynamics

Bernardo García Bulle Bueno[1], Salome Aguilar Llanes[2], Tobin South[1], Alex 'Sandy' Pentland[1], and Esteban Moro[1,3]

[1]**Connection Science, MIT. Cambridge, MA**
[2]**Economics, MIT. Cambridge, MA**
[3]**Universidad Carlos III, Madrid**

ABSTRACT

Understanding the mechanisms of effective education is of paramount importance to society, and an area of research in need of more quantitative analysis. Much of the effectiveness of education settings lies in the nuance of the two way communication in the learning processes, and insights into how this communication meditates learning is hidden in natural language text. To understand more about the process, we performed an RCT on a tutoring program where university students teach Math to groups of up to five elementary school kids in 2021. We randomized both the tutor-group matching, as well as the focus of the class: just on Math or both Math and socioemotional learning. Using recordings of the sessions, we asked whether latent variables obtained from the scripts of the lessons could reveal (a) the mean increase in Math scores of the students in the group before and after participating in the program, and (b) the treatment that was assigned. We find that latent features from the script predict those metrics with 12 and 25% r-squared respectively. Math improvement was better predicted using latent variables from earlier layers of an encoder, while the opposite was true for the content. This suggests that identifying whether socioemotional content was part of the class was more related to the structure or meaning of the sentences, while Math improvement was more related to the words used.

Keywords: RCT, Education, NLP

1 INTRODUCTION

We propose to use tools from natural language processing to analyse the dynamics of learning within the classroom environment. To achieve this, we will use a new dataset collected through a tutoring RCT in Mexico comprising several thousands of hours of Math tutoring sessions with associated information on student and teacher outcomes in a Math standardized test implemented before and after tutoring session began. The tutoring connected groups of 5 elementary school kids to each tutor, and kids could be assigned to just Math content, or Math+socioemotional learning content. All the sessions were recorded and then analyzed by extracting latent variables from a Spanish language encoder. Automating the task of "observing" classroom dynamics and teaching style with a consistent toolkit allows us to understand how the detailed dynamics of classrooms affect outcomes at scale.

1.1 Related work

On the side of classroom observation the project conducted by the gates foundation "Measures of Effective Teaching" has a similar setting to ours as in that case classes were recorded and recordings were analyzed by observers using a standard set of questions that helped build an index of the measure. Students were randomly assigned to teachers within each school. The researchers observed that teachers that initially got a higher score of the measure were likely to get it again even if the students were changed to a new set of random students within the school [3]. There have been efforts to develop an internationally valid instrument to measure teacher effectiveness [5].

There are several examples of the use of Natural Language Processing (NLP) tools in education for instance the "E-rater" developed by ETS or Criterion to predict essay scores. Other existing tools such as "Text Adaptor" that help teachers develop text adaptations for their students. [2]. NLP can also aid in e-learning. For instance "Language Muse" that can give linguistic feedback to students. NLP can help in the analysis of linguistic errors and aid as a tool to the teacher [1]. There also have been applications of NLP trying to replicate one-to-one human tutoring interactions. As well as to process web material to personalize instructional materials [4].

1.2 Data

In this work we analyze whether the script of a class can be used to predict either the Math improvement of the kids in that class, or whether the class covered only Math topics, or Math along socioemotional learning activities as well. To do it we passed the recordings of classes through an automatized speech-to-text software which yielded a text file containing the dialogue that was said in the class in a written form. Due to computational and budget constraints, only classes for 45 groups were analyzed in the following section.

2 REPRESENTATION LEARNING OF LATENT VARIABLES

To obtain latent variables from the scripts, we used an existing encoder architecture trained to understand Spanish. This approach is better than training a model on our own given our dataset is small compared to one used to train a language model. For this we used nvidia's Spanish English translator (24x6) with 24 encoding layers and 6 decoding layers. The architecture is a transformer one and includes key, query and value representations in the tensor. We only focused on the encoder part as that's the part that provides us with representations of the Spanish language.

We considered the intermediate product between layers of the encoder to be our representation. For each layer we analyzed, we extracted five variables: the mean of first n vectors (corresponding to n words), the last word vector (last word), and the 3 vectors corresponding to query, key and value, separately.

2.1 Relationship to outcome variables

We decided to use two outcome variables per group: Average improvement of kids in Math tests, and Assignment to emotional+Math tutoring (compared to just Math). Then we did two regressions explaining each of those outcome variables (y) with the latent variables we found (x). We use Lasso regularized OLS in both cases. To evaluate the performance of each we did k-fold out-of-sample predictions for each datapoint, and measured performance in terms of r-squared.

3 RESULTS

For this section we obtained latent variables from an the encoder of a Spanish-English translator as described in 2. We tested features from five parameters: mean word embedding, last word embedding, and query, key and value features from the transformers. Moreover, we also had the option to compare the results across different layers. The results are plotted in figure 1.

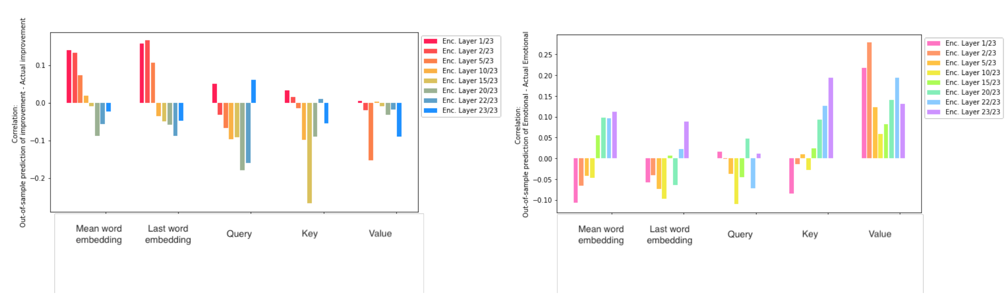

Figure 1. Results for different layers in terms of correlation between ground truth and out of sample k-fold predictions. The bar height is the median correlation across different regularization parameters.

We obtain very optimistic results in the representation learning. Curiously, for Math outcomes, the best results are found using early layer word encodings. As in other domains, the deeper intermediate layers are the worst, with an uptick in performance from the last layer. For the emotional treatment layer, the words are not as effective, and only are in the last layers. However, the value layer is surprisingly effective at capturing the emotional treatment class.

This result shows that encoded in the sentences of the class we have information which is related to the success and format of the class. The next step in our case was to be able to interpret the factors correlated with each of the outcome variables.

To be able to do it we used the original transformer model. We took a random subset of all sentences in all classes. Then for each sentence we obtained the values of the layer which we want to interpret. To do it, we generated predictions for the sentences and sorted them by the outcome variable to see how they differed.

In the case of the emotional treatment in the class, we found that longer sentences were associated with those classes. That is coherent with previous findings we observed which were that students felt more confident to talk in class in that treatment. This opens the door to being able to characterise a class by the emotional content of it and the bonding which happened in it.

Unfortunately, while we do see predictive power for characterising classes with more Math improvement, the methods we tried did not yield any results which we could qualitatively interpret with clarity, for which they are omitted.

4 DISCUSSION & CONCLUSION

This analysis has leveraged NLP to advance our knowledge of how education happens within the classroom. We find that despite the limitations of current pretrained models on Mexican Spanish, the use of multilingual models makes transcription effective and mostly robust. Utilising a learned representation space from pretrained translation models allowed for strong predictive performance of education outcomes with limited explainability. To address this we identify characteristic sentences that match high importance characteristics of the representation space and qualitatively identify that emotion and human connection in the educational context is a strong determinate of learning outcomes.

Our approach has the potential to help us understand how kids learn and can be used to help teachers improve their skills and learn from others. In future work, we propose to link more outcome variables to features encoded in language models. One important area of improvement we identified was our strategy to interpret the variables, as it was purely qualitative and relied on small samples of sentences. We hope to improve this by using automatic text generation techniques, and other visualisation techniques which can shed light on which exact components of the sentence resulted in higher perceived emotional content in the class.

REFERENCES

[1] Khaled M Alhawiti. "Natural language processing and its use in education". In: *Computer Science Department, Faculty of Computers and Information technology, Tabuk University, Tabuk, Saudi Arabia* (2014).

[2] Jill Burstein. "Opportunities for natural language processing research in education". In: *International Conference on Intelligent Text Processing and Computational Linguistics*. Springer. 2009, pp. 6–27.

[3] Thomas J Kane et al. "Have we identified effective teachers? Validating measures of effective teaching using random assignment". In: *Research Paper. MET Project. Bill & Melinda Gates Foundation*. Citeseer. 2013.

[4] Diane Litman. "Natural language processing for enhancing teaching and learning". In: *Thirtieth AAAI conference on artificial intelligence*. 2016.

[5] Charles Teddlie et al. "The international system for teacher observation and feedback: Evolution of an international study of teacher effectiveness constructs". In: *Educational research and evaluation* 12.6 (2006), pp. 561–582.