Learning Average Course Load Hours from Student Responses

By Rujul Singh, Kitty Moy, Nicholas Romig, Noelle Jung
CSCI 1470 (Deep Learning), Department of Computer Science, Brown University

BROWN

Introduction

Course load is one of the most important factors that college students take into consideration when carving out their schedules for the school year. At Brown University, there is a considerable amount of data submitted to Critical Review thanks to the students who take the time to respond. However, many classes still lack the numerical data of course load compared to other types of data like answers to open-ended questions and professor ratings. In our project, we aim to solve this issue by employing a combination of Natural Language Processing, or NLP, numerical data analysis, and regression. We attempt to predict average course load hours per week with three different models that take disparate sets input data.

Methodology

Data

The dataset is actual student responses for the Critical Review of various undergraduate classes from the past five years. There are 2844 data points in total with course load hours per week ranging from zero to thirty-five, text answers to open-response questions, and many more numerical answers.

Models

We implemented three different RNN model architectures: Base RNN Model, RNN with Dropout, and Multi-Input RNN integrated with Numerical Data. They were multi-channel multi-input models that incorporated Natural Language Processing and numerous connected layers but varied in the inputs fed. Losses were calculated via mean squared error.

Architecture

Base RNN Model

The input for this model was text answers to the following three questions:
1. What types of assignments did this course have? Please check the appropriate boxes and provide the number/frequency/any other relevant information about each type of course work
2. Is there anything else prospective students should know about this course?
3. Discuss the instructor's teaching style. What was effective and what was not?

The responses were tokenized and lifted into a 100-dimensional embedding space.

Three separate Gated Recurrent Units were used to account for the different contexts provided by the three questions. The three sets of resultant vectors from the GRUs were each processed through two dense layers with ReLu activation and were concatenated into one set. Lastly, it was processed through a single dense layer with ReLu activation and a regression dense layer to yield a final output that represents the model's predicted "average hours per week" for the course.

RNN with Dropout

This model was identical to the base RNN model but included a dropout layer in the embedding matrix.

Multi-Input RNN Integrated with Numerical Data

All attributes from the dataset, including numerical answers to fifteen fields and natural language data were combined to create an integrated end-to-end regressor. The numerical inputs processed through dense layers were concatenated with the three textual data processed with the GRUs, and finally processed with a regression layer

Results

The model achieved significant accuracy after being trained. Random guessing of average hours from zero to thirty-five hours, yields a MSE of over 200. The results from our three models with all the hyperparameters - epochs, batch size, learning rate - tuned to their optimal values are in the table below.

Model	MSE
RNN	14.92
RNN with Dropout	14.96
Multi-input Model	14.38

Incorporating dropout in the model did not appear to show any improvements to the accuracy. The addition of numerical data led to a slight improvement of 0.6. These results suggest that textual data from the open-response questions are sufficient indicators of average workload.

Challenges

We implemented parallel pipelines of neural nets, which have not been encountered in class, in order to process the mixed data types and incorporate them altogether into a single output. There were also problems of punctuation and capitalization inconsistencies and spelling errors.
The original dataset was too small, so we had to increase the size of our dataset to just under 3000 points from the past five years in the Critical Review archive.
The model was very sensitive to hyperparameter choices. The same learning rate and number of epochs from the first two models were not yielding great results for the third model. It was also computationally intensive given the amount of hyperparameters, so it took the model many hours to try to find the optimal set of parameters using a grid-based search pattern.

Discussion

The resulting mean square errors are significantly lower than just randomly guessing the number of hours. An error of approximately 3.8 hours could be overlooked considering it is for over a week. Although our mean square error did not reach our base, target, and stretch goals of less than 5, 3, and 1 hour, respectively, we predict that the results could improve with more data points. Additionally, we had to compromise with the hyperparameters to keep a consistent baseline for the three models, concerned about drastically changing them from method to method, hence the small improvement by the third model despite the incorporation of numerical data. Further hyperparameter optimization could improve the MSEs more.

For future purposes, the first improvement to make to the project is increasing the size of the dataset, which was a limiting factor to our model with just under 3000 data points. Unfortunately, we already exhausted the Critical Review's digitized dataset. Getting more data would be outside of the scope of this project where translating scans of old reviews would require immense hours of manually inputting data or an entirely new deep learning algorithm to digitize the texts. Another improvement would be to build a model with just a numerical dataset to separate its predictive capacity from the text data in the third model.

Incomplete data was a major problem we encountered in this project, but courses that have significant text data can still have accurate predictions for the number or hours required per week, which would allow students to more easily plan their schedules. We believe that our model creates the opportunity for much more information about potential classes to be communicated to students.