CS410 Project Proposal - ArguMentor

Felipe Assumpção, Rafael Piacsek

October 2023

1 Introduction

1.1 Team members:

- Felipe Assumpção (*fza2*, lead),
- Rafael Piacsek (*rafaelp5*, member).

1.2 Topic:

1.2.1 Project Overview:

Argumentative writing is an essential skill for academic and professional success. However, a significant portion of underprivileged high school students struggle with proficient writing. One effective way to address this issue is by building and deploying an automated, instant feedback tool capable of evaluating students' writing and providing personalized feedback. This project aims to create a viable solution that enhances the quality of automated writing evaluation, making it more accessible to educators and students.

1.2.2 Project Objectives:

- **Text Segmentation**: Develop algorithms to automatically segment texts in essays written by $6^{th} - 12^{th}$ grade students. Accurate segmentation is crucial for correctly and consistently identifying the distinct rhetorical and argumentative elements in a piece of writing. This is where we plan to either make use of MetaPy or potentially make use of other NLP strategies with other Python packages (PyTorch?) to achieve segmentation.
- **Argumentative and Rhetorical Element Classification:** With the segmented text, we then create models to classify such argumentative and rhetorical elements within students' essays. Accurate identification of these elements is key for providing effective targeted feedback. This part will most likely make use of PyTorch to develop a neural NLP classification model.

- **Open-Source Solution:** Open-source the developed algorithms to make them widely accessible to educational organizations free of cost, enabling them to leverage this project to improve the writing skills of young students.
- **User-Friendly Interface (Stretch goal):** Develop a user-friendly interface for easy uploading of essays and provide clear, actionable feedback.

1.2.3 Expected Outcomes:

- **Improved Writing Proficiency:** By creating a tool that effectively segments and classifies students' essays into rhetorical and argumentative essays, we expect that this tool can be incorporated into other education tools for enhanced and immediate feedback. Therefore, we aim to aid students in improving their writing skills, as more targeted and personalized feedback can lead to better writing outcomes.
- **Wider Accessibility:** By open-sourcing the algorithms we develop for this project, we hope to assist educational institutions with limited resources to benefit from this technology and, in turn, contribute to reducing disparities in rhetorical and argumentative writing proficiency.
- **Project Evaluation:** As briefly discussed below, we will split the dataset between training, validation, and testing. Thus, we will have labeled, novel data to benchmark our final solution against.

1.3 Programming language and frameworks:

- Python, PyTorch, MetaPy(?)

1.4 Workload:

- Dataset (1.5) Analysis + Preprocessing: 5 hours
- Base NLP Model for Segmentation: 10 hours
- Segment Classification Model: 10 hours
- Optimize/Improve Models: 20 hours
- Command Line Interface, Project Wrap-Up: 5 hours
- (*Stretch goal*) Graphical User Interface or Website: 20 hours

1.5 Resources:

- Dataset: Kaggle
- Evaluation: The dataset (1.5) itself will be split into training, validation, and testing sections.