HW3 Report

John Pascoe

Github: [CPSC-8430-Deep-Learning/HW3 at main · jpascoe32fb/CPSC-8430-Deep-Learning (github.com)]

Introduction

The ability to accurately extract answers from spoken content, despite noise and variability in speech, is a significant challenge in natural language processing (NLP). This project aimed to address this challenge by leveraging a BERT model, trained on the Spoken SQuAD dataset, to perform extractive question answering. By evaluating the model's performance across different levels of noise variance, my research seeked to understand the resilience and adaptability of BERT in extracting answers from noisy spoken text.

Methodology

Data Preparation

The Spoken SQuAD dataset served as the foundation for training. This dataset comprises spoken versions of the SQuAD texts, providing a unique challenge due to its inclusion of speech-specific nuances. The testing datasets were prepared with three varying levels of added noise, simulating real-world audio environments.

Model Configuration

The project utilized a BERT model for its robustness and effectiveness in NLP tasks. Modifications specific to spoken language processing and noise resilience were applied. The model was fine-tuned with the following strategies:

- **Tokenization**: Utilized BERT's tokenizer to handle spoken language peculiarities, ensuring the model could effectively process speech-derived text.
- **Window Splitting**: Implemented to manage longer texts, dividing them into manageable pieces for the BERT model to process efficiently.
- **Learning Rate Scheduling**: A linear decay learning rate scheduler was employed to optimize the training process, adjusting the learning rate over time for better convergence.
- **Gradient Accumulation**: Used to handle hardware limitations, allowing the model to effectively learn from larger batches of data by accumulating gradients over multiple steps before updating model parameters.

Training

The model was trained on the Spoken SQuAD dataset, with particular attention to optimizing for the nuances of spoken text and the added noise conditions. Training details, including epochs, batch size, and specific hardware used, are recorded for readability within the .ipynb file for this project. They were designed to be reproducible and to provide insights into the training dynamics of the model under the specific conditions.

Evaluation

The model's performance was evaluated on the test sets with varying levels of noise variance. The F1 score metric was used to quantify the model's ability to accurately extract answers from noisy spoken content. Additional analyses on the impact of noise levels on model performance were conducted to understand how different types of noise and their intensities affect the overall task. There were slight variations of scores for the three test types, with the accuracy of my model generally decreasing as the noise increased.

Conclusion

This project demonstrates that a BERT model, when carefully fine-tuned and tested, shows promise in extracting answers from noisy spoken content. My findings suggest that with appropriate training strategies, including window splitting and gradient accumulation, and careful adjustment of learning rates and postprocessing, models like BERT can be adapted to handle the complexities of spoken language question and answering. However, performance variations across different noise levels highlight the ongoing challenge in dealing with real-world audio environments. Future work could explore more sophisticated noise reduction techniques and further adapt models for better resilience against various types of auditory disturbances.