Project Proposal

Enhancing Question-Answering Systems with BERT

Ayushi Chakrabarty
Georgia Institute of Technology
achakrabarty8@gatech.edu

Cameron Potter
Georgia Institute of Technology
cpotter8@gatech.edu

Kshitij Pathania
Georgia Institute of Technology
kpathania3@gatech.edu

Prateek
Georgia Institute of Technology
p34@gatech.edu

Sneha Maheshwari
Georgia Institute of Technology
smaheshwari63@gatech.edu

Project Group 50

Introduction/Background

In the evolving landscape of Natural Language Processing (NLP), the quest for models that can understand and generate human-like responses has become paramount. Among the myriad of models that have emerged, BERT (Bidirectional Encoder Representations from Transformers) stands out as a revolutionary architecture that has set new benchmarks in a range of NLP tasks. Specifically, in the domain of Question-Answering (QA) systems, the adaptability and prowess of BERT can lead to significant enhancements in accuracy, contextual understanding, and response generation. This work delves deep into the nuances of integrating BERT into QA tasks, showcasing its potential to reshape the way machines understand and answer questions, bringing them a step closer to human-like conversational abilities.

Problem Definition

The primary challenge is to investigate and evaluate how BERT can be effectively integrated into existing or novel QA frameworks to enhance their performance. This includes:

1. How can BERT's bidirectional understanding of context be harnessed to improve the accuracy of QA systems?

2. What modifications or fine-tuning techniques are required to adapt BERT specifically for diverse QA tasks, considering the variability in question types and domains?

3. How can the scalability and efficiency of QA systems be maintained or improved upon integrating the computationally intensive BERT model?

4. In what ways can BERT's capabilities be leveraged to make QA systems more robust against ambiguous, misleading, or poorly framed questions?

Methods

To optimize QA systems using BERT, our strategy encompasses multiple stages. Initially, we'll tap into prominent QA datasets like CommonsenseQA, and SQuAD to curate a broad spectrum of questions, subsequently tailoring this data for BERT's prerequisites. Building on this foundation, we'll integrate the pre-trained BERT model and finetune it to cater specifically to QA demands. The performance of our adapted model will be scrutinized based on key indicators, such as prediction accuracy, answer quality gauged via the F1 score, and responsiveness. To further refine the model and curb overfitting, we'll delve into regularization and optimization avenues. Ensuring the model's robustness is crucial; hence, it will be assessed against both generic and complex queries. Recognizing BERT's intricate architecture, efforts will be directed to demystify its decision-making, enhancing its interpretability. In the deployment phase, our vision is to create a scalable, efficient, and widely accessible model. Overall, this holistic approach is designed to harness BERT's prowess, amplifying the QA system's precision, transparency, and operational efficiency.

Potential Results and Discussion

As we integrate BERT into QA systems, we aim to enhance prediction accuracy, response efficiency, robustness against diverse queries, model interpretability, and scalability. Expected results include increased accuracy and F1 scores, highlighting BERT's strengths in language comprehension. However, a potential trade-off may arise between model complexity and responsiveness due to BERT's computational demands. Our research will also probe BERT's decision-making layers and assess the model's adaptability for large-scale deployment. Key metrics for evaluation comprise prediction accuracy, F1 score, response latency, and coverage of relevant answers. Ultimately, our findings will provide insights into BERT's capabilities in QA, guiding future NLP innovations.

1 Proposed Timeline

The timeline is divided in three phases across the whole semester. Our streamlined timeline is set to effectively harness BERT for QA tasks within the stipulated period. The link to the Gantt chart is mentioned here: Gantt Chart

References

1. Rajpurkar, P., Zhang, J., Lopyrev, K., & Liang, P. (2016). SQuAD: 100,000+ Questions for Machine Comprehension of Text. https://huggingface.co/datasets/squad

2. Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding.

3. Vaswani, A., Shazeer, N., Parmar, N., et al. (2017). Attention Is All You Need.

4. Talmor, A., & Berant, J. (2019). CommonsenseQA: A Question Answering Challenge Targeting Commonsense Knowledge. https://www.kaggle.com/datasets/thedevastator/new-commonsenseqa-dataset-for-multiple-choice-qu

5. Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep learning.

Contribution Table

Student Name	**Contributed Aspects**
Ayushi Chakrabarty	Potential Results and Discussions, Gantt chart, Video Creation
Cameron George Potter	Ideation of the project theme, Methods, Video Creation
Kshitij Pathania	Introduction and Background, Documentation, Video Creation
Prateek	Potential Dataset, Proposal presentation, Video Creation
Sneha Maheshwari	Problem Definition, Documentation, Video Creation

Table 1: Contributions of team members

Checkpoint

As of the proposal submission, we have procured and initialized datasets, including SQuAD, and CommonsenseQA to drive our research on integrating BERT into QA systems. While we are committed to our initial dataset choices, the dynamic nature of research might necessitate changes. Should there be a need to switch or modify our dataset, a comprehensive explanation will be provided, detailing why the initial dataset was deemed unsuitable in the upcoming reports.