A Question Answering NLP System for Graduate Advising using BERT
Bailleddygbar, Sivasankarreddy (MS Electrical Engineering)
Kondaparthi, Priyanka (MS Electrical Engineering)

Introduction
Natural Language Processing (NLP) is the ability of machines to understand and interpret human language. Sentiment Analysis, Next Sentence Prediction and Question Answering (QA) are the popular applications of NLP. The prominence of Question Answering systems is growing in customer-centric businesses where customer service is the focus. The success of QA systems is due to the introduction of state-of-the-art BERT (Bidirectional Encoder Representations from Transformers) [1] language model and SQuAD (Stanford Question Answering Dataset) [2]. BERT is trained on Wikipedia text and interprets each word in relation to sentence context with human level accuracy.

The project aims to develop an automated Question Answering system that acts as a graduate advisor and responds to students’ queries. In contrast to traditional QA systems, the project proposes a novel approach where users query the system without entering the input context. The final deliverable of the project is a web application with an interface that allows students to enter question, select category and view the predicted model response. The dataset for the project is collected with the help of graduate advisors and FAQs section of department website. The project uses BERT finetuned on SQuAD as baseline model. The project employs Jaccard similarity score as the performance metric.

Dataset
The project uses Natural Questions (NQ) dataset [3] containing google search queries, data collected with the help of project advisor and Electrical Engineering department website for finetuning the BERT base model. The training dataset size is around 50K. The dataset contains question, paragraph and start and end positions. Figure 1 show the distribution of NQ dataset.

Methodology

BERT Model
BERT base model shown in Figure 2 contains 12 hidden layers with 768 states each. BERT architecture is a stack of encoders. Transformers architecture which enables attention mechanism [4] is the heart of BERT. Attention mechanism allows BERT to maintain context between the words far apart in a sentence.

Model Fine-tuning
The training dataset is converted to JSON format before feeding to the BERT model for fine-tuning. BERT uses Word Piece tokenization to convert the question and context to tokens. BERT vocabulary file and embedding lookup table converts each input token to a vector of length 768 which is hidden state size. The maximum input sequence length is 512 tokens. BERT outputs two sequences containing probabilities for each token being the start token and end token.

Analysis and Results
The application database contains tables for storing context text, categories and log history. The application server handles communication between the user and the API. This promotes modularity and allows the model to be modified without any concern about the application.

Table 1. Jaccard score of fine-tuned model

<table>
<thead>
<tr>
<th>Question</th>
<th>Predicted Answer</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>When shall I remove my admission conditions?</td>
<td>You should complete all admission conditions stated in your acceptance letter as soon as possible to avoid the delay of your graduation.</td>
<td>58.8%</td>
</tr>
<tr>
<td>Do I have to take 2 core courses during the first semester at SJSU?</td>
<td>Most of the elective courses require backgrounds from the core courses, completing the 2 core courses at the end of your first semester will be an advantage for you.</td>
<td>23.3%</td>
</tr>
</tbody>
</table>

Table 2. Comparison of results

<table>
<thead>
<tr>
<th>Question</th>
<th>Predicted Answer</th>
<th>Baseline</th>
<th>Fine-tuned</th>
</tr>
</thead>
<tbody>
<tr>
<td>When shall I remove my admission conditions?</td>
<td>You should complete all admission conditions stated in your acceptance letter as soon as possible to avoid the delay of your graduation.</td>
<td>0%</td>
<td>58.8%</td>
</tr>
<tr>
<td>Do I have to take 2 core courses during the first semester at SJSU?</td>
<td>Most of the elective courses require backgrounds from the core courses, completing the 2 core courses at the end of your first semester will be an advantage for you.</td>
<td>7.69%</td>
<td>23.3%</td>
</tr>
</tbody>
</table>

Summary & Conclusions
Natural Language Processing provides opportunities to develop useful application across various industries. An automated QA system that can respond to academic related queries is beneficial to students. The project outcome is a web application that allows students to enter the query and provides the response. Data gathering, finetuning the BERT model, developing and deploying the inference API are major project tasks. The Jaccard similarity scores of the fine-tuned model are higher than those of the base model which is not fine-tuned. The future scope is expansion to a conversational model.

Key References

Acknowledgements
We would like to express sincere gratitude to our project advisor Dr. Birsen Sirkeci for her constant support and guidance. We are thankful to her for pointing us towards the right resources and more importantly for providing good learning experience.

We are thankful to Electrical Engineering department for their hard work and continuous support to students.