NLP Assignment - 3 Report

Abhirama Subramanyam
penamakuri.1@iitj.ac.in

Kanika Choudhary
choudhary.10@iitj.ac.in

May 21, 2020

The task is to setup a fine-tune a BERT QA System, using the covid-19 question and paragraph data (of 205 Questions). Here we present the results from our custom data fine-tuned BERT model and compare it with original BERT model as well as SQUAD v1.1 fine-tuned model. Clearly, our model has given better results than both of them. We also present some examples to showcase the same.

1 Dataset

1. **For custom fine tuning** - We have **manually** created a small dataset of covid-19 (Question/Answers/Context) pairs from the shared csv file containing (question/context) pairs. Our dataset is in accordance with Standard **SQUAD v1.1** format. We have split the data into two parts, 150 examples for fine-tuning (for convenience, we refer this as train data) and 55 examples as test data.
2. **For SQUAD Fine-tuning** - SQUAD v1.1

2 Our Models

1. BERT-Small (L-4,H-512) (No Fine-tuning).
2. BERT Fine-tuned on SQUAD v1.1 data.
3. BERT Fine-tuned on SQUAD v1.1 data followed by Fine-tuning on Custom Data.

3 Training Details

We have pulled up the pre-trained BERT from official google's BERT GitHub repository. We've tried training BERT-Basic, BERT-Medium, but due to system constraints, i.e., they needed lot of compute power resources (at least 16GB Graphics Card), So we picked up BERT-small which has 4 Layers with 512 hidden parameters.

1. learning rate = 3e-5
2. train_batch_size = 6/12
3. epochs = 2

Model - 1, pre-trained BERT-Small (L-4, H-512) model.
Model - 2, was fine-tuned from Model - 1 checkpoint using SQUAD v1.1 data.
Model - 3 was fine-tuned from Model - 2 checkpoint using custom train data.

4 Files

- JSON_150_updated.json - custom data train file.
- JSON_50_updated.json - custom data test file.
- train-v1.1.json - SQUAD 1.1 train file.
- dev-v1.1.json - SQUAD 1.1 test file.
- Comparison report (of test questions on 3 models).
- process_json.py - Snippet to clean process JSON.
- print_results.py - File to give comparison results.
- model-50.ckpt - Model-3 (Fine-tuning on SQUAD followed by Custom Data)
- model-29199.ckpt - Model-2 (Fine-tuned on SQUAD alone).
- custom_finetuned_results.json - Results by Custom Fine-tuned BERT.
- original_bert.json - Results by Original BERT.
- squad_finetuned_results - Results by SQUAD v1.1 Fine-tuned BERT.

*Models and SQUAD v1.1 are not attached, as they are heavy, will be shown during demo.

5 Results

The following are the results obtained from different techniques. It is clearly evident that our model has given better results than other two on questions (55) which are unseen during fine-tuning. Some of the examples are listed below.

ID	**152**
Question	Is it okay for me to donate blood ?
Original Answer	people who are well to continue to donate blood
Result by Original BERT (No Fine-tuning)	19, but it is likely since this is the case with other viruses like flu. caring for patients
Result by SQUAD 1.1 Fine-tuned BERT	constant
Result by CUSTOM DATA Fine-tuned BERT	The need for donated blood is constant, and blood centers are open and in urgent need of donations

ID	**157**
Question	how do I prepare my children in case of covid-19 outbreak in our community ?
Original Answer	Talk with your children about the outbreak, try to stay calm, and reassure them that they are safe
Result by Original BERT (No Fine-tuning)	19. Cough or sneeze into your elbow or use a tissue. If you use a tissue
Result by SQUAD 1.1 Fine-tuned BERT	mild
Result by CUSTOM DATA Fine-tuned BERT	Talk with your children about the outbreak, try to stay calm, and reassure them that they are safe

ID	**158**
Question	How can my family and I prepare for Covid-19 ?
Original Answer	Wash your hands with soap and water for at least 20 seconds. Stay one metre or more away from people
Result by Original BERT (No Fine-tuning)	19 outbreak occurs in your community
Result by SQUAD 1.1 Fine-tuned BERT	use your garden
Result by CUSTOM DATA Fine-tuned BERT	Create an emergency contact list of family, friends, neighbors, carpool drivers, health care providers, teachers, employers, the local public health department

ID	**162**
Question	Should I use soap and water or hand sanitizer to protect against covid-19 ?
Original Answer	soap and water are not readily available, use an alcohol-based hand sanitizer
Result by Original BERT (No Fine-tuning)	near-real-time surveillance (observation) of all
Result by SQUAD 1.1 Fine-tuned BERT	If soap and water are not readily available, use an alcohol-based hand
Result by CUSTOM DATA Fine-tuned BERT	Wash your hands often with soap and water for at least 20 seconds, especially after blowing your nose, coughing, or sneezing

ID	**171**
Question	Are there any vaccines or other medical products to prevent covid-19 ?
Original Answer	no vaccine
Result by Original BERT (No Fine-tuning)	19, he/she will inform you of the procedure to follow and advise where and how the test
Result by SQUAD 1.1 Fine-tuned BERT	vaccines, antibodies, and drugs
Result by CUSTOM DATA Fine-tuned BERT	no vaccine

ID	**175**
Question	Are antibiotics effective in preventing or treating the COVID-19?
Original Answer	No
Result by Original BERT (No Fine-tuning)	complications than any other healthy individuals
Result by SQUAD 1.1 Fine-tuned BERT	Antibiotics do not prevent
Result by CUSTOM DATA Fine-tuned BERT	Antibiotics do not prevent or treat coronavirus disease (COVID-19), because COVID-19 is caused by a virus, not bacteria

ID	**177**
Question	How are people tested for covid-19 ?
Original Answer	sample is typically collected from your nose and/or throat with a special swab
Result by Original BERT (No Fine-tuning)	19, a sample is typically collected from your nose and/or throat with a special swab at a designated collection location
Result by SQUAD 1.1 Fine-tuned BERT	a sample is typically collected from your nose and/or throat
Result by CUSTOM DATA Fine-tuned BERT	a sample is typically collected from your nose and/or throat with a special swab at a designated collection location staffed by health care professionals

ID	**179**
Question	Can animals get covid-19 disease ?
Original Answer	no scientific evidence
Result by Original BERT (No Fine-tuning)	fusion-transmitted coronavirus. Flexible
Result by SQUAD 1.1 Fine-tuned BERT	can spread COVID-19 or that they might be a source of infection in the United States
Result by CUSTOM DATA Fine-tuned BERT	no evidence that companion animals, including pets, can spread COVID-19 or that they might be a source of infection in the United States

ID	**184**
Question	How does COVID-19 spread
Original Answer	can spread from person to person through small droplets from the nose or mouth which are spread when a person with COVID-19 coughs or exhales
Result by Original BERT (No Fine-tuning)	near someone with corona virus or lived in or traveled to an area
Result by SQUAD 1.1 Fine-tuned BERT	coughs or exhales
Result by CUSTOM DATA Fine-tuned BERT	small droplets from the nose or mouth which are spread when a person with COVID-19 coughs or exhales

ID	**185**
Question	Can the virus that causes COVID-19 be transmitted through the air?
Original Answer	through contact with respiratory droplets rather than through the air
Result by Original BERT (No Fine-tuning)	hand hygiene after touching or discarding the mask Use alcohol-based hand rub or, if visibly soiled, wash your hands with soap and water
Result by SQUAD 1.1 Fine-tuned BERT	through contact with respiratory droplets
Result by CUSTOM DATA Fine-tuned BERT	through contact with respiratory droplets rather than through the air

ID	**188**
Question	What can I do to protect myself and prevent the spread of disease ?
Original Answer	You can reduce your chances of being infected or spreading COVID19 by taking some simple precautions
Result by Original BERT (No Fine-tuning)	-19 hotspots (cities or local areas where COVID-19 is spreading widely). If possible
Result by SQUAD 1.1 Fine-tuned BERT	reduce your chances of being infected or spreading COVID19 by taking some simple precautions
Result by CUSTOM DATA Fine-tuned BERT	avoid traveling to places especially if you are an older person or have diabetes, heart or lung disease

ID	**192**
Question	Are antibiotics effective in preventing or treating the COVID-19?
Original Answer	No
Result by Original BERT (No Fine-tuning)	-19 outbreak occurs in your community
Result by SQUAD 1.1 Fine-tuned BERT	COVID-19 is caused by a virus
Result by CUSTOM DATA Fine-tuned BERT	antibiotics do not work. Antibiotics should not be used as a means of prevention or treatment

ID	**193**
Question	Are there any medicines or therapies that can prevent or cure COVID-19 ?
Original Answer	no evidence that current medicine can prevent or cure the disease
Result by Original BERT (No Fine-tuning)	-19 hotspots (cities or local areas where COVID-19 is spreading widely). If possible
Result by SQUAD 1.1 Fine-tuned BERT	no evidence that current medicine can prevent or cure the disease. We does not recommend self-medication with any medicines
Result by CUSTOM DATA Fine-tuned BERT	no evidence that current medicine can prevent or cure the disease. We does not recommend self-medication with any medicines

ID	**194**
Question	Is there a vaccine drug or treatment for COVID-19 ?
Original Answer	no vaccine and no specific antiviral medicine
Result by Original BERT (No Fine-tuning)	and young
Result by SQUAD 1.1 Fine-tuned BERT	Possible vaccines and some specific drug treatments
Result by CUSTOM DATA Fine-tuned BERT	no

ID	**199**
Question	Can I catch COVID-19 from my pet??
Original Answer	no evidence that a dog, cat or any pet can transmit COVID-19
Result by Original BERT (No Fine-tuning)	and even death
Result by SQUAD 1.1 Fine-tuned BERT	mainly spread through droplets produced when an infected person coughs, sneezes, or speaks.
Result by CUSTOM DATA Fine-tuned BERT	No

ID	**203**
Question	What is the government doing to contain the spread of COVID-19?
Original Answer	following the instructions of the WHO carefully and every level of public health machinery is aligned to it
Result by Original BERT (No Fine-tuning)	-19 by touching these objects or surfaces
Result by SQUAD 1.1 Fine-tuned BERT	following the instructions of the WHO carefully and every level of public health machinery is aligned to it
Result by CUSTOM DATA Fine-tuned BERT	following the instructions of the WHO carefully and every level of public health machinery is aligned to it

ID	**205**
Question	How long is the incubation period for COVID-19?
Original Answer	range from 1-14 days, most commonly around five days
Result by Original BERT (No Fine-tuning)	affected geographic areas. Community spread means people have been infected with the virus in an area
Result by SQUAD 1.1 Fine-tuned BERT	five days
Result by CUSTOM DATA Fine-tuned BERT	1-14 days, most commonly around five days

6 References

1. Google BERT
2. SQUAD v1.1
3. Compact BERT Models