

# BERT

BERT for NLP has taken the AI by storm since its launch in 2018. In 2019, Google announced that it had begun using BERT in its search engine, and by the end of 2020, BERT will be used in almost all English searches. BERT is characterized by considering the context or the difference in the meaning of words appearing in different sentences in the learning process, so it has become one of the most popular NLP algorithms.

# Applications

- Use BERT to answer sentence questions, extract fixed feature vectors, and analyze sentence semantics.

# How to use

BERT APP is divided into three parts:

- SQuAD2.0 conducts sentence question training and answer
- ELMo extracts fixed feature vectors
- MRPC conducts sentence semantic analysis.

The screenshot displays the LEADERG - BERT - Version 6 application interface, which is divided into three main sections: MRPC, SQuAD2.0, and ELMo.

**MRPC Section (Blue Border):**

- 1. Fine\_Tuning:** Includes a "batch size" input field set to 16.
- 2. Inference:** Includes a "model file" input field with the path "data/uncased\_L-12\_H-768\_A-12/bert\_model.ckpt" and a "Choose" button.
- View execution input file:** data/glue-data/MRPC/test.tsv
- View execution output file:** data/output/MRPC/test\_results.tsv

**ELMo Section (Green Border):**

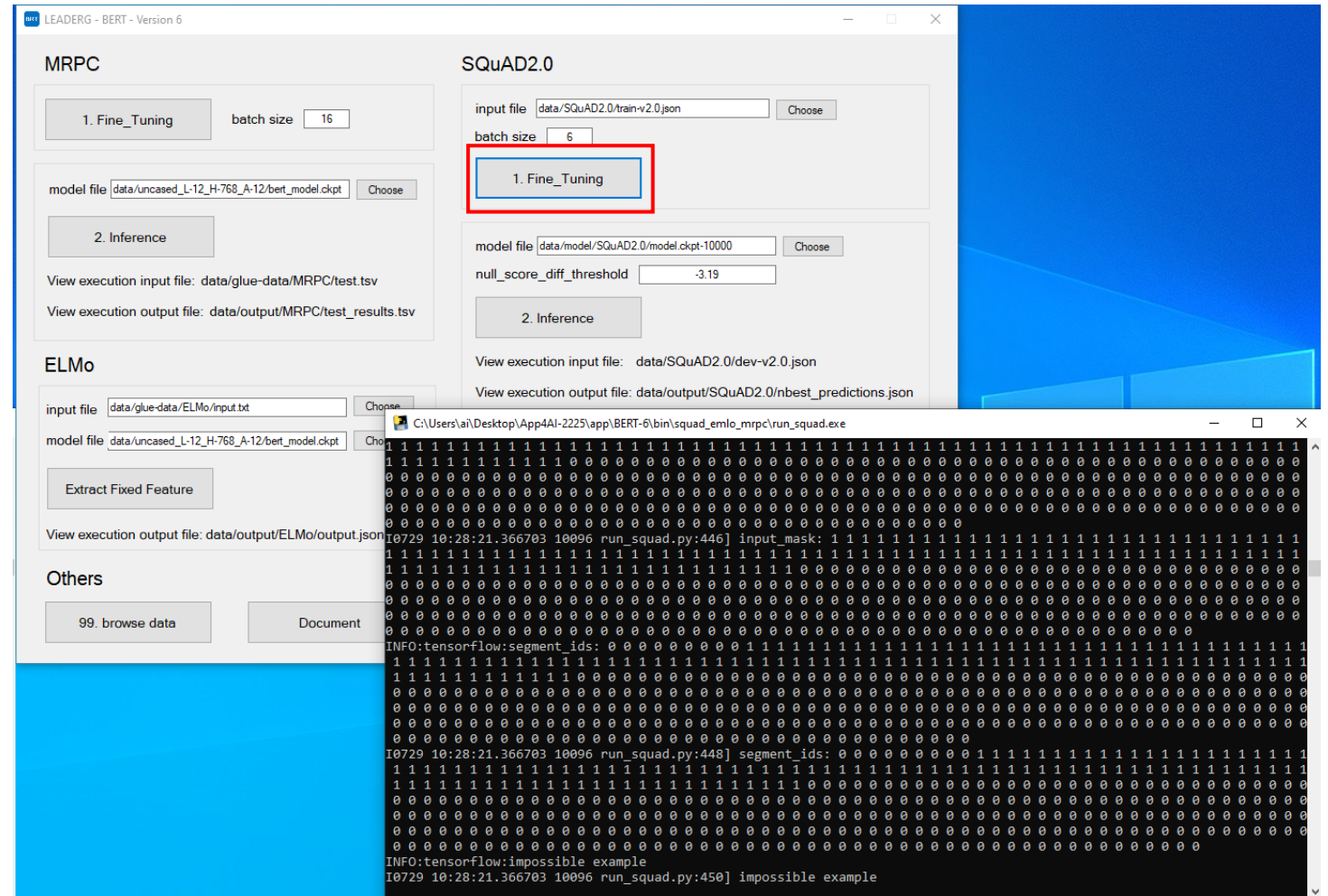
- Input file:** data/glue-data/ELMo/input.txt (with "Choose" button)
- Model file:** data/uncased\_L-12\_H-768\_A-12/bert\_model.ckpt (with "Choose" button)
- Extract Fixed Feature:** A button to initiate feature extraction.
- View execution output file:** data/output/ELMo/output.json
- Others:** Includes a "99. browse data" button and a "Document" button.

**SQuAD2.0 Section (Yellow Border):**

- 1. Fine\_Tuning:** Includes an "input file" input field with the path "data/SQuAD2.0/train-v2.0.json" and a "Choose" button. The "batch size" input field is set to 6.
- 2. Inference:** Includes a "model file" input field with the path "data/model/SQuAD2.0/model.ckpt-10000" and a "Choose" button. The "null\_score\_diff\_threshold" input field is set to -3.19.
- View execution input file:** data/SQuAD2.0/dev-v2.0.json
- View execution output file:** data/output/SQuAD2.0/nbest\_predictions.json
- 3. Evaluate:** Includes a button to initiate evaluation.
- View execution input file:** data/SQuAD2.0/dev-v2.0.json
- View execution output file:** data/output/SQuAD2.0/predictions.json
- View execution output file:** data/output/SQuAD2.0/null\_odds.json

# SQuAD2.0 - 1. Fine\_Tuning

- In the input file of the SQuAD2.0 area, select the file data/SQuAD2.0/train-v2.0.json to be fine-tuned, and press 1. Fine\_Tuning will start the fine-tuning.
- After pressing it, it will stop and wait for a while in the "impossible example" line, which is a normal phenomenon.

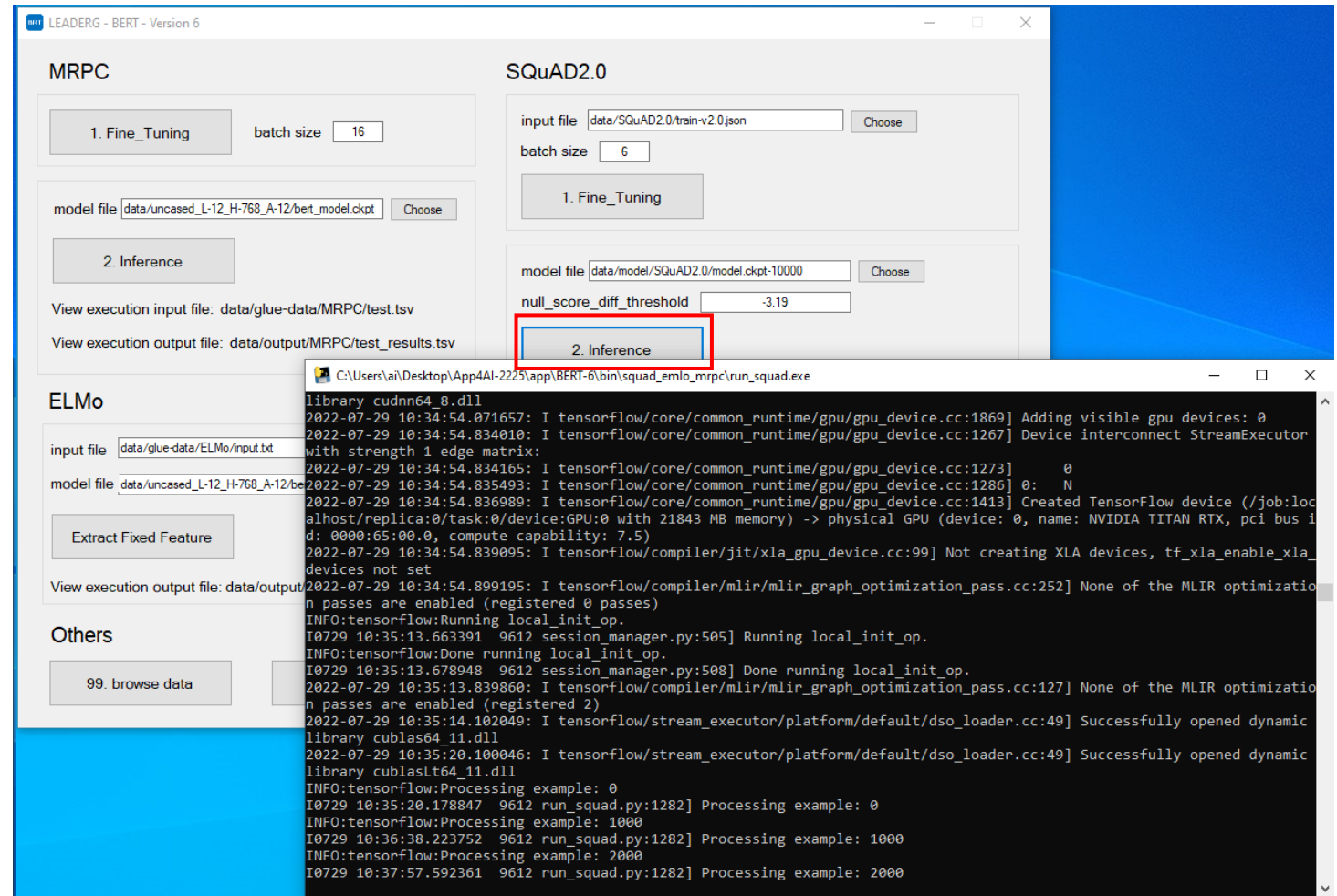


# SQuAD2.0 - 2. Inference

After training, you can choose the training model to answer the sentence question. The question file is from data/SQuAD2.0/dev-v2.0.json, and the output answer is in data/output/SQuAD2.0/nbest\_predictions.json. If there is no answer to the question, the threshold of the question will be recorded in data/output/SQuAD2.0/null\_odds.json.

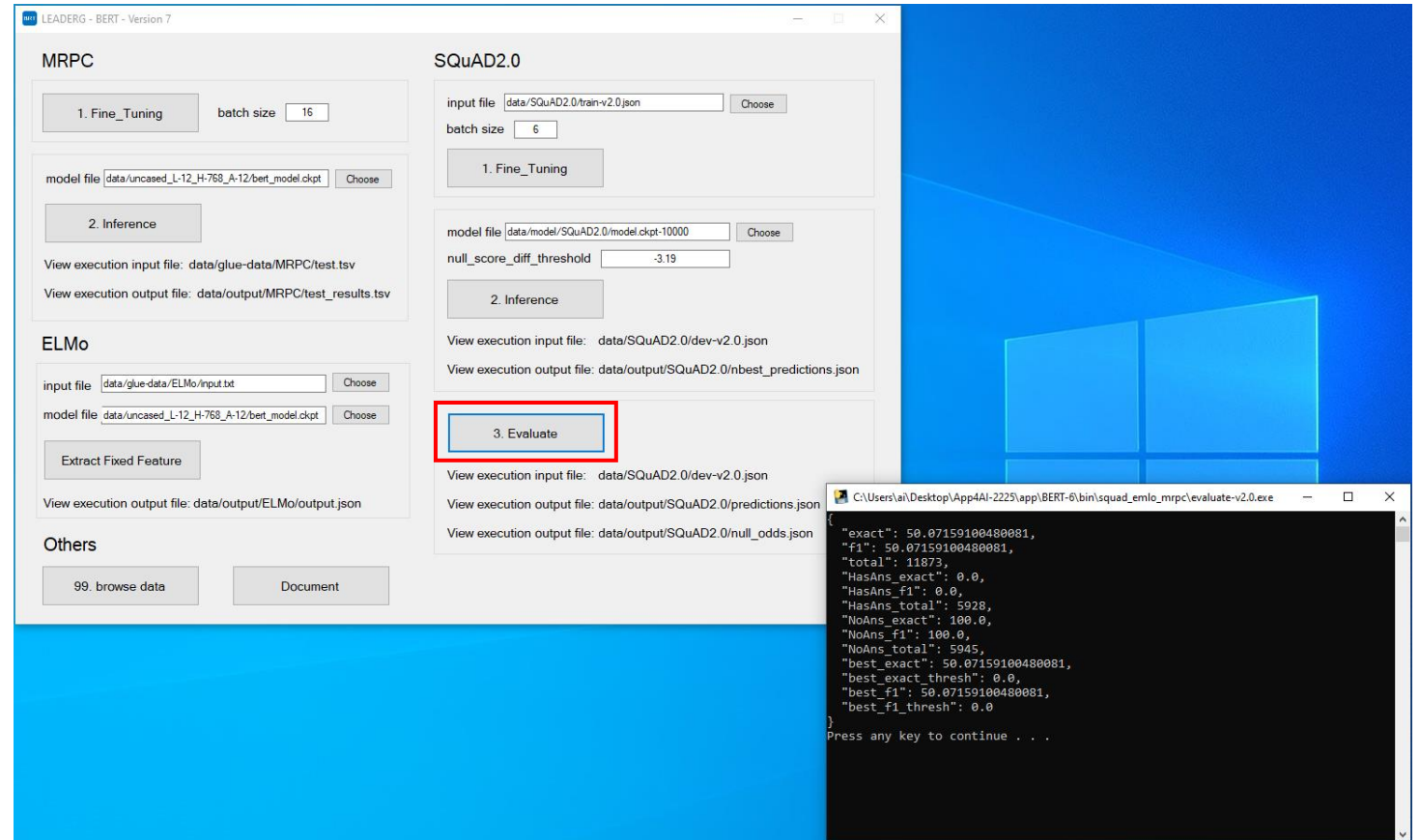
Note:

When you select model file, three files, model.ckpt-XXX.data-00000-of-00001, model.ckpt-XXX.index, and model.ckpt-XXX.meta, will be read. Do not delete the extension data and. index file.



# SQuAD2.0 - 3. Evaluate

If you need to adjust the threshold of the inference answer, you can execute "3. Evaluate" and adjust the threshold of null\_score\_diff\_threshold according to the value of best\_f1\_thresh in the execution result.



# SQuAD2.0 – Inference File Description

The position of the orange line in nbest\_predictions.json corresponds to the id of the question in the source file data/SQuAD2.0/dev-v2.0.json, and the answer to the question is in [ ].

dev-v2.0.json - Notepad

File Edit Format View Help

```
{
  "version": "v2.0",
  "data": [
    {
      "title": "Normans",
      "paragraphs": [
        {
          "text": "The Normans were a people who in the 10th and 11th centuries gave their name to Normandy in France. They were descended from Norsemen who invaded England from Scandinavia. They were also known as the Normans, and the name was derived from the word 'Norman' which means 'from the north' or 'from the northmen'."
        }
      ],
      "questions": [
        {
          "text": "When were the Normans in Normandy?",
          "id": "56ddde6b9a695914005b9628",
          "answers": [
            {
              "text": "France",
              "answer_start": 159
            }
          ],
          "is_impossible": false
        },
        {
          "text": "From which country did the Normans first gain their independence?",
          "id": "56ddde6b9a695914005b962a",
          "answers": [
            {
              "text": "Denmark, Iceland and Norway",
              "answer_start": 256
            }
          ],
          "is_impossible": false
        },
        {
          "text": "What century did the Normans first gain their independence?",
          "id": "56ddde6b9a695914005b962b",
          "answers": [
            {
              "text": "10th century",
              "answer_start": 649
            }
          ],
          "is_impossible": false
        },
        {
          "text": "Who did King Charles III swear fealty to?",
          "id": "56ddde6b9a695914005b962c",
          "answers": [
            {
              "text": "Normandy",
              "answer_start": 137
            }
          ],
          "is_impossible": true
        },
        {
          "text": "Who did King Charles III swear fealty to?",
          "id": "56ddde6b9a695914005b962d",
          "answers": [
            {
              "text": "Normandy",
              "answer_start": 137
            }
          ],
          "is_impossible": true
        }
      ]
    }
  ]
}
```

nbest\_predictions.json - Notepad (Not Responding)

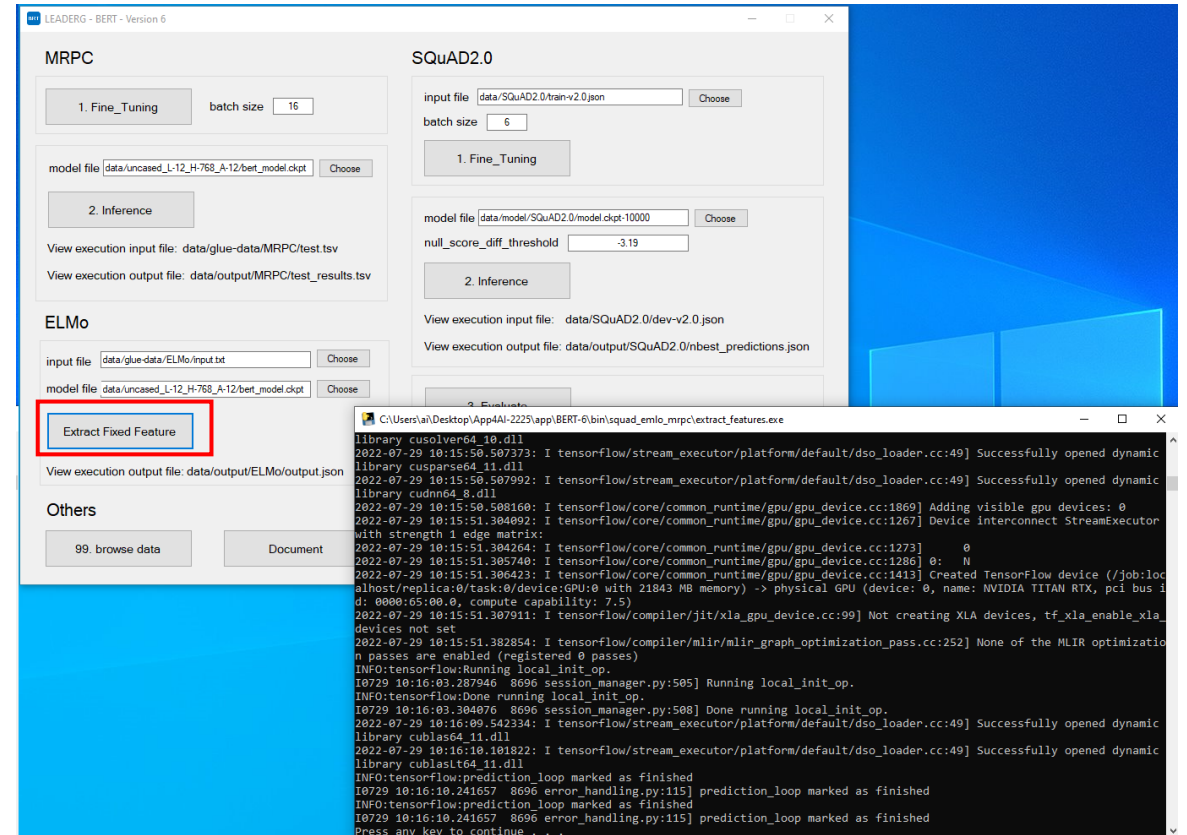
File Edit Format View Help

```
{
  "56ddde6b9a695914005b9628": [
    {
      "text": "",
      "probability": 0.9730640992296529,
      "start_logit": 6.8147687911987305,
      "end_logit": 7.538949489593506
    },
    {
      "text": "French: Normands; Latin: Normanni) were the people who in the 10th and 11th centuries gave their name to Normandy in France. They were descended from Norsemen who invaded England from Scandinavia. They were also known as the Normans, and the name was derived from the word 'Norman' which means 'from the north' or 'from the northmen'."
    },
    {
      "text": "10th and 11th",
      "probability": 0.0017323133921425611,
      "start_logit": 3.7598485946655273,
      "end_logit": 4.262877464294434
    }
  ],
  "56ddde6b9a695914005b962a": [
    {
      "text": "Denmark, Iceland and Norway",
      "probability": 0.0017323133921425611,
      "start_logit": 3.7598485946655273,
      "end_logit": 4.262877464294434
    }
  ],
  "56ddde6b9a695914005b962b": [
    {
      "text": "10th century",
      "probability": 0.0017323133921425611,
      "start_logit": 3.7598485946655273,
      "end_logit": 4.262877464294434
    }
  ],
  "56ddde6b9a695914005b962c": [
    {
      "text": "Normandy",
      "probability": 0.0017323133921425611,
      "start_logit": 3.7598485946655273,
      "end_logit": 4.262877464294434
    }
  ],
  "56ddde6b9a695914005b962d": [
    {
      "text": "Normandy",
      "probability": 0.0017323133921425611,
      "start_logit": 3.7598485946655273,
      "end_logit": 4.262877464294434
    }
  ]
}
```



# ELMo – Extract Fixed Feature

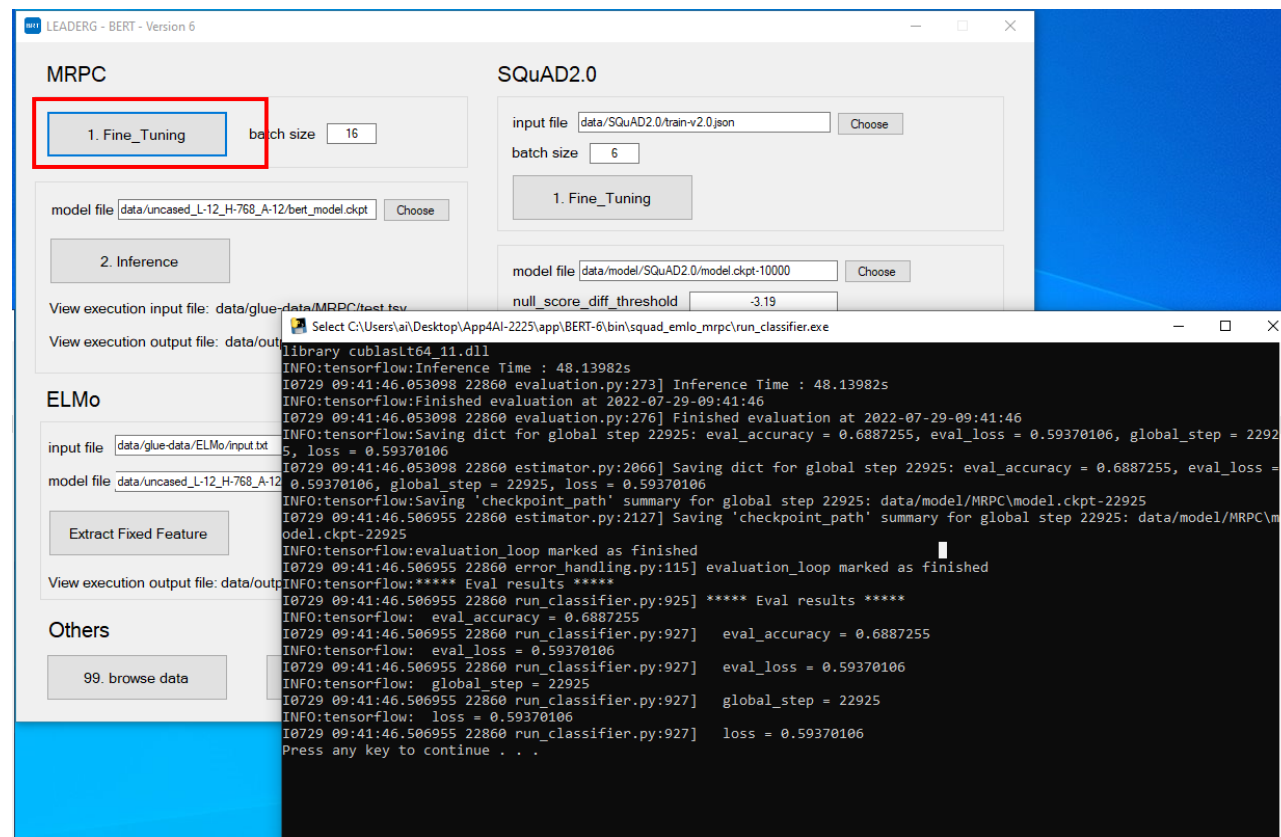
Press "Extract Fixed Feature" to extract the fixed feature vector from input file “data/glue-data/ELMo/input.txt”, and store the extraction result in “data/output/ELMo/output.json”.





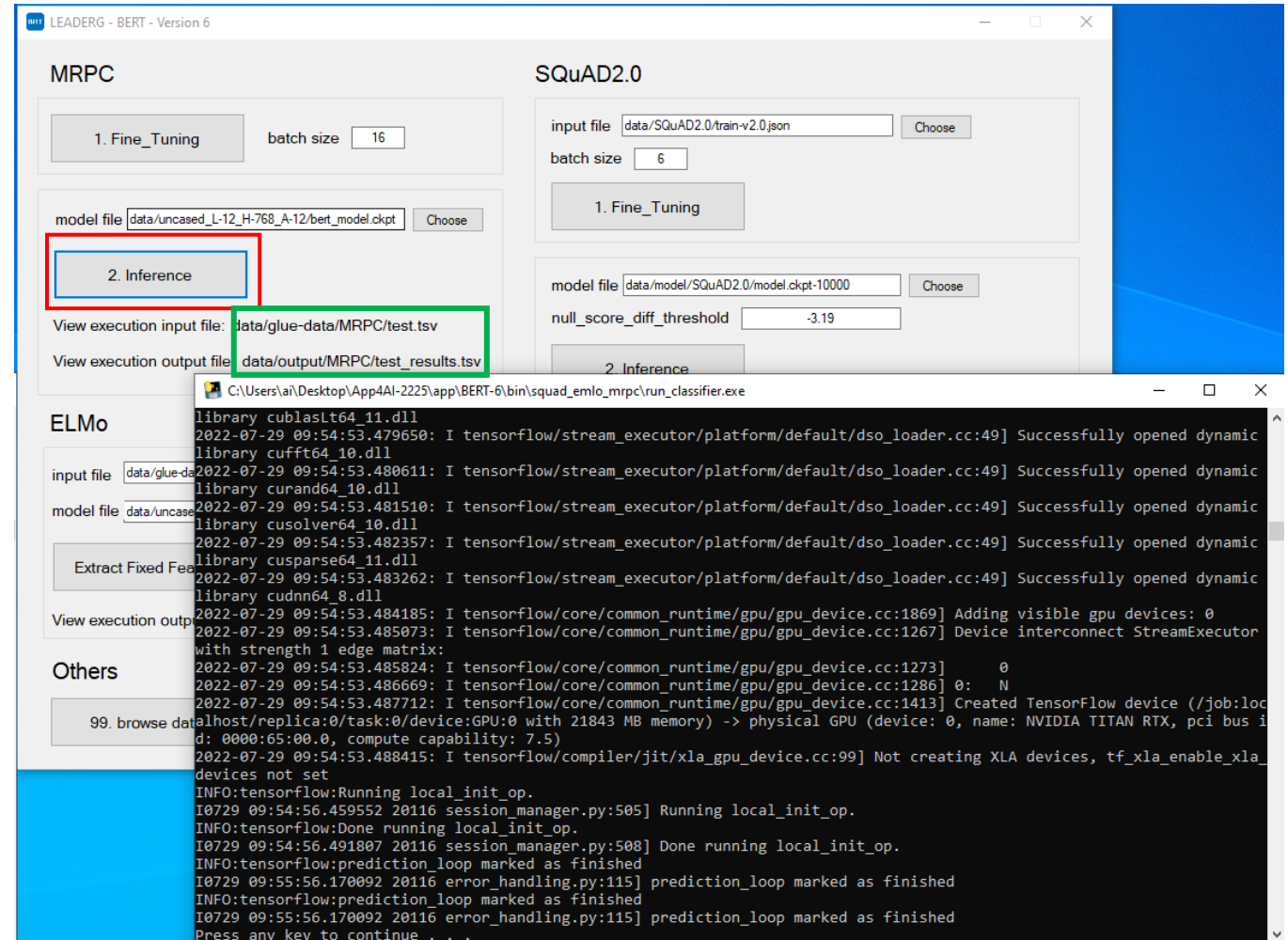
# MRPC - 1. Fine\_Tuning

After pressing "1. Fine\_Tuning",  
train the files in the “data/glue-  
data/MRPC” folder.



# MRPC - 2. Inference

- After training, confirm the model file and press "2.Inference" to perform sentence semantic analysis on the "data/glue-data/MRPC/test.tsv" file to determine the probability that the two sentences belong to the same meaning.
- Click the tsv path in the green box to view the inference source file and the inference result file.



# MRPC – Inference File Description

The green line of test.tsv is sentence 1, and the blue line is sentence 2. After analysis, at the orange line of test\_results.tsv, the former number (0.3868928) indicates the probability that the two sentences have different meanings, and the latter number (0.6131072) indicates The probability that two sentences have the same meaning.

test.tsv - Notepad					test_results.tsv - Notepad				
File Edit Format View Help					File Edit Format View Help				
index	#1 ID	#2 ID	#1 String	#2 String	0.3868928	0.6131072			
0	1089874	1089925	PCCW 's chief operating officer , Mike Butcher , and Alex Arena , the chief financial officer , will report directly to Mr So . Current Chief Operating Officer Mike Butcher and Group Chief Financial Officer Alex Arena will report to So .		0.3589465	0.6410535			
1	3019446	3019327	The world 's two largest automakers said their U.S. sales declined more than predicted last month as a late summer sales frenzy caused more of an industry backlash than expected . Domestic sales at both GM and No. 2 Ford Motor Co. declined more than predicted as a late summer sales frenzy prompted a larger-than-expected industry backlash .		0.45561427	0.5443857			
2	1945605	1945824	According to the federal Centers for Disease Control and Prevention ( news - web sites ) , there were 19 reported cases of measles in the United States in 2002 . The Centers for Disease Control and Prevention said there were 19 reported cases of measles in the United States in 2002 .		0.35719386	0.6428062			
3	1430402	1430329	A tropical storm rapidly developed in the Gulf of Mexico Sunday and was expected to hit somewhere along the Texas or Louisiana coasts by Monday night . A tropical storm rapidly developed in the Gulf of Mexico on Sunday and could have hurricane-force winds when it hits land somewhere along the Louisiana coast Monday night .		0.45730287	0.54269713			
4	3354381	3354396	The company didn 't detail the costs of the replacement and repairs . But company officials expect the costs of the replacement work to run into the millions of dollars .		0.3648189	0.63518107			
5	1390995	1391183	The settling companies would also assign their possible claims against the underwriters to the investor plaintiffs , he added . Under the agreement , the settling companies will also assign their potential claims against the underwriters to the investors , he added .		0.46577832	0.5342217			
6	2201401	2201285	Air Commodore Quaife said the Hornets remained on three-minute alert throughout the operation . Air Commodore John Quaife said the security operation was unprecedented .		0.44386667	0.55613333			
					0.5479067	0.45209327			
					0.37929586	0.6207042			
					0.45919302	0.540807			
					0.3549672	0.6450328			
					0.37342593	0.6265741			
					0.6291312	0.37086874			
					0.35217297	0.647827			
					0.38554677	0.61445326			
					0.42968297	0.57031703			
					0.5081551	0.49184495			
					0.35947794	0.64052206			
					0.35333169	0.64666826			
					0.4979651	0.5020349			
					0.40600004	0.59300001			

# References

- Please refer to the readme.txt in the APP folder.
- LEADERG AppForAI: <https://www.leaderg.com/appforai-windows>
- Copyright © LEADERG INC. All rights reserved.