Fine Tuning of MiniLM for Lie Detection Task

Project By :
- Cveevska Marija
- Erim Suleyman
- Karakus Isikay
- Varagnolo Mattia

Mentor :
- Prof. Giuseppe Sartori
- Dr. Merylin Monaro

Table of contents

Introduction

Humans are not very good at telling when someone is lying
and human accuracy in detecting deception is very constrained.
This study delves into the fine-tuning of MiniLM Language Model,
exploring the efficacy in discerning deceit and lie detection.

LLMs,known as Transformer language models, pack in hundreds of millions
of parameters and are trained on vast collections of texts in their initial phase.
This pre-training helps LLMs understand the intricate structures of the language.

Once they have gone through this pre-training phase,
these models can be tuned further for specific jobs using smaller sets of data.
We will compare fine-tuning methods on already pre trained LLM and
also do different scenarios concerning the data set for lie detection

01

Objectives of the project

Objectives

Our aim

The main objective of this project is to delve into LLMs fine-tuning and compare different methods and different scenarios for Lie Detection.

The goal

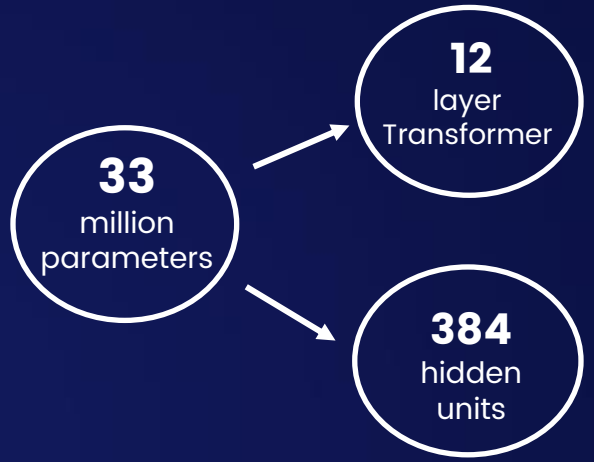
Accuracy Improvement: Enhancing the model's accuracy in distinguishing between truthful and deceitful statements.

02

MiniLM – Pre Trained Large Language Model

MiniLM - LLM

Description



- **MiniLM** is a task-agnostic, compressed pre-trained transformer model. Achieves a notable compression factor of 2.7x compared to BERT-Base.
- Its multilingual variant, maintains compatibility **with XLM-R's tokenizer with BERT's Transformer** architecture. It is equipped with **21 million Transformer parameters and 96 million embedding parameters.** Demonstrates robust and competitive performance on cross-lingual tasks.
- For English tasks, MiniLM offers several pre-trained models , these models use the same WordPiece vocabulary as BERT.

03

Dataset – Future Intentions

Dataset – Future Intentions

Description

- The Intention dataset, focusing on Future Intentions, consists of 1640 statements which focuses on participants' most significant nonwork-related activities. The dataset has six columns with capturing participants' responses to two key questions:

- Q1: "Please describe your activity as specifically as possible."
- Q2: "Which information can you provide to reassure us that you are telling the truth?"

- The dataset also features an "outcome class" column indicating whether each statement is truthful (T) or deceptive (F). It encompasses 857 deceptive and 783 truthful statements, with each participant providing two answers.

04

Fine Tuning Scenarios

Parameter-Efficient Fine-Tuning (PEFT) with LoRA

Parameter-Efficient Fine-Tuning (PEFT) with LoRA (Layer-wise Relevance Adjustment) is a technique designed to improve the fine-tuning process in transfer learning, particularly in scenarios where there's a need to adapt a pre-trained model to a new task while efficiently utilizing parameters.

Parameter-Efficient Fine-Tuning (PEFT): PEFT controls the fine-tuning process to prevent overfitting and enhance parameter efficiency.

LoRA (Layer-wise Relevance Adjustment): It adds additionally weights on top of pretrained ones, and in this way, it does not cause catastrophic forgetting like in full fine tuning.

Transfer Learning

Transfer learning is a technique where a model trained on one task is reused or adapted as a starting point for a new but related task.
Essentially, the model leverages its pre-existing knowledge to enhance efficiency in learning new tasks. In the context of fine-tuning MiniLM, we incorporate transfer learning through a structured process encompassing three key steps:
the pre-training phase,
the fine-tuning phase, and
the inference phase.
Freezing layers during transfer learning involves keeping certain layers of a pre-trained model fixed (not updating their weights) while allowing others to be trained on new data.

Scenarios

Two Scenarios:

Scenario 1 - Transfer Learning
□ MiniLM with Fine Tuning Method – Transfer Learning
□ Full Dataset with freezing of 0, 3, 4 and 9 layers
□ Intentions statements - Q1 , Q2 and Q1 + Q2

Scenario 2: Parameter-Efficient Fine-Tuning (PEFT) with LoRA
□ MiniLM with Fine Tuning Method – PEFT with LoRA
□ Small Dataset and Full Dataset
□ Intentions statements - Q1 , Q2 and Q1 + Q2

05

Results

Results: Scenario 1 - Transfer Learning

Dataset	Question type	Epochs	Val Accuracy	Test Accuracy	Layers frozen
Full	q1+q2	30	77.50%	70.00%	4
Full	q2	30	57%	59.76%	4
Full	q1	30	70%	69.51%	4
Full	q1	30	62%	65.85%	3
Full	q2	30	65.85%	64.63%	3
Full	q1+q2	30	64.63%	62.20%	3
Full	q1	30	62.20%	64.64%	9
Full	q2	30	52.44%	53.66%	9
Full	q1+q2	30	64.63%	67.07%	9
Full	q1	30	66%	67.07%	0
Full	q2	30	62%	59.75%	0
Full	q3	30	59.75%	67.07%	0

Results :Scenario 2: Parameter-Efficient Fine-Tuning (PEFT) with LoRA

Dataset	Question Type	Epochs	Val Accuracy	Test Accuracy
Full	q1+q2	50	56.098%	59.76%
Full	q1+q2	30	56.098%	59.76%
Full	q1+q2	10	56.098%	59.76%
Full	q2	50	56.098%	59.76%
Full	q2	30	56.098%	59.76%
Full	q2	10	56.098%	59.76%
Full	q1	50	56.098%	59.76%
Full	q1	30	56.098%	54.88%
Full	q1	10	56.098%	59.76%
800 samples	q1+q2	50	52.500%	57.50%
800 samples	q1+q2	30	52.500%	57.50%
800 samples	q1+q2	10	52.500%	57.50%
800 samples	q2	50	52.500%	57.50%
800 samples	q2	30	70.000%	65.00%
800 samples	q2	10	52.500%	57.50%
800 samples	q1	50	67.500%	52.50%
800 samples	q1	30	75.000%	65.00%
800 samples	q1	10	52.500%	57.50%

06

Conclusion

Conclusion

When comparing the two Fine Tuning methods we can see that the Transfer Learning method shows best results in general.

Despite our efforts in hyperparameter tuning and model selection, we acknowledge that our achieved maximum accuracy of 70\% may not meet the desired threshold for optimal performance.

For further improvements in model performance, experiments can be conducted with more extensive configurations along with the search for superior optimization algorithms and improved learning rates. This can enable additional insights from the data to optimize convergence and strike a better balance between training efficiency and accuracy.

Thank you

For your attention

Click the icon for the full project