Assignment2 Report

Xuan Chen - #82286956

In this assignment, I trained and evaluated the baseline model and the knowledge base model using trainer.predict, based on COPA validation dataset (including 100 samples).

Error Type

According to the example of error analysis, I categorized the top prediction into the following categories:

U1: The prediction has no relationship with premise.

P2: The prediction is correct somehow, but not 100% correct.

E3: The prediction is an effect, but the question requires a cause.

R4: The relationship between prediction and premise is not effect and cause.

C5: The prediction contradicts the premise.

From my observation, I don't think there is any linguistic property that is shared between some of the errors.

I think common sense knowledge (BERT may not have a deep understanding of the real-world situations and common sense knowledge that humans have acquired through experience and observation) and cultural and social knowledge (BERT may not fully understand the cultural and social contexts in which language is used, which can affect the meaning and interpretation of language) are not expected to have in BERT.

Performance of the model

```
{'1': 32, '2': 20, '3': 8, '4': 12, '5': 28}
{'1': 32, '2': 24, '3': 16, '4': 12, '5': 16}
```

	Validation Accuracy %	Error Distribution %				
		1	2	3	4	5
BERT	70	32	20	8	12	28
BERT with KB	61	32	24	16	12	16
probability			0.7	0.15		0.18

Table. Evaluation results of two models

```
1: stat, probability: TtestResult(statistic=0.0, pvalue=1.0, df=99)
2: stat, probability: TtestResult(statistic=-0.3763388118272598, pvalue=0.7074703580131823, df=99)
3: stat, probability: TtestResult(statistic=-1.4214106244380287, pvalue=0.15833990565972564, df=99)
4: stat, probability: TtestResult(statistic=0.0, pvalue=1.0, df=99)
5: stat, probability: TtestResult(statistic=1.3470946333202294, pvalue=0.18102514023295704, df=99)
```

From the table, we can see that the performance of the model declines after add Knowledge base. For example:

"[CLS] patient is people. people has information. malpractice is like physician. malpractice is like patient. effect is in the context of physic. physic is like physician. the physician misdiagnosed the patient. what happened as a result? [SEP] the patient disclosed confidential information to the physician. [SEP]']"

It's probably because:
- Increased complexity: Incorporating a knowledge base increases the complexity of the input, which can make it more difficult for the model to process and understand the information. This can result in lower accuracy and slower inference times.
- Conflicting information: The knowledge base may contain information that contradicts the information in the input text. This can cause confusion for the model, leading to incorrect predictions and lower performance.
- Limited or biased knowledge: The knowledge base may be limited in scope or biased towards certain perspectives, which can result in incomplete or inaccurate understanding of the input.
- Structural differences: The structure and format of the knowledge base may differ from the natural language used in the input, which can make it harder for the model to integrate and use the knowledge effectively.

Incorporating a knowledge base can improve the model's performance by providing additional information and context that is relevant to the task at hand. For example, a knowledge base can help BERT understand domain-specific terminology or facts that are not commonly found in text.