CPSC 532V: Common sense Reasoning in NLP
Assignment #2

Sadaf Sadeghian

Step 2) Train a Neural Model

For this step I changed the pre-processing part of the notebook and after training the model, it achieved the loss mentioned below:

Epoch	Training Loss	Validation Loss	Accuracy
1	0.657600	0.691006	0.520000
2	0.672200	0.678567	0.570000
3	0.630800	0.673610	0.600000

Then 25 examples that the model predicted incorrectly were sampled which can be categorized in the following categories:

Errors:

1) Ambiguity (28%)

Examples:

```
Premise: My eyes became red and puffy.
Question: cause
  Choice1 - I was sobbing.
  Choice2 - I was laughing.
```

Ambiguity - Someone might tear up when they laugh.

```
Premise: The man dressed in his best suit.
Question: cause
  Choice1 - He scheduled a meeting with an important client.
  Choice2 - His wife bought him a new tie.
```

Ambiguity – He might also want to wear the new tie with his best suit.

```
Premise: I put my hands under the running faucet.
Question: effect
```

```
  Choice1 - The soap rinsed off my hands.
  Choice2 - The water splashed in my face.
```

Ambiguity – When you put your hand under the water, if the water pressure is not adjusted it might splash to someone's face.

```
Premise: My skin broke out into a rash.
Question: cause
  Choice1 - I brushed against poison ivy in my yard.
  Choice2 - I eradicated the poison ivy from my yard.
```

Ambiguity – For eradicating the poison ivy you might need to touch it which may cause a rash

2) Lack of Knowledge about Objects (knowledge about properties of something) (20%) [let's call it Factual Knowledge]

```
Premise: I pressed my hand into the wet cement.
Question: effect
  Choice1 - My handprint dried in the cement.
  Choice2 - Cracks emerged in the cement.
```

Lack of Factual Knowledge – About the properties of wet cement

```
Premise: The flame on the candle went out.
Question: cause
  Choice1 - I blew on the wick.
  Choice2 - I put a match to the wick.
```

Lack of Factual Knowledge – How the candle goes out.

3) Lack of Knowledge about behaviors of human (52%) [let's call it Common-sense Knowledge]

```
Premise: The woman covered her mouth with her hand.
Question: cause
  Choice1 - She exhaled.
  Choice2 - She sneezed.
```

Lack of common-sense Knowledge – About the appropriate behaviours.

```
Premise: The woman resigned from her job.
Question: cause
  Choice1 - She aspired to hold an executive position in the firm.
  Choice2 - She believed her superiors were acting unethically.
```

Lack of common-sense Knowledge – When do people resign?

```
Premise: The judge pounded the gavel.
Question: cause
  Choice1 - The courtroom broke into uproar.
  Choice2 - The jury announced its verdict.
```

Lack of common-sense Knowledge – When do judge pounds the gavel?

Step 3) Train a Neuro-Symbolic Model

For this step, I used the text-based approach to incorporate the knowledge from the knowledgebase to the model. Then to choose between multiple paths between the word pairs, I assigned a score to each path which is the sum of the weights of the edges in that path normalized by the length of the path. Then I picked the top 3 paths with the highest path score as the most relevant paths. These context sentences (the natural language representation of the paths) are then added to the premise and fed into the BERT.

By adding the common-sense data to the model, the performance improved in comparison to the purely neural model. The metrics are shown below:

Epoch	Training Loss	Validation Loss	Accuracy
1	0.703800	0.686324	0.530000
2	0.590200	0.685398	0.660000
3	0.434000	0.719998	0.670000

By observing the performance of the model, we can conclude that the model could deduct better using the additional common-sense data provided to it. Mostly, in the examples which needed some kind of knowledge which BERT didn't have (factual or common-sense).

Then 25 examples that the model predicted incorrectly were sampled which can be categorized in the following categories:

Errors:

1) Ambiguity (24%)

Some examples are ambiguous themselves, which even having more knowledge cannot help the

```
Premise: The hamburger meat browned.
Question: cause
  Choice1 - The cook froze it.
  Choice2 - The cook grilled it.
```

Ambiguity – The frozen meet is also browned sometimes when it's oxidised and expires.

2) Misleading Symbolic knowledge (76%)

The knowledge we added was harmful instead of helpful because it might show irrelevant data to the sentences or even find better matches for the wrong answer and worse context for the correct answer which causes the model to predict the wrong answer.

```
Premise: The man drank heavily at the party.
Question: effect
  Choice1 - He had a headache the next day.
  Choice2 - He had a runny nose the next day.
```

Misleading Symbolic knowledge – Knowledge added seems to connect the wrong answer better with the premise than the correct answer.

(Choice1: party is used for birthday. birthday is related to day. man and person are have similar meanings. person wants headache. drank is a form of the word drink. person does not want drink. person wants headache.
Choice2: party is used for birthday. birthday is related to day. drank is a form of the word drink. drink is related to liquid. liquid is similar to runny. eye is related to party. eye is related to nose.)

```
Premise: I wanted to conserve energy.
Question: effect
  Choice1 - I swept the floor in the unoccupied room.
  Choice2 - I shut off the light in the unoccupied room.
```

Misleading Symbolic knowledge – Knowledge added was not helpful

('energy is related to electricity. You are likely to find electricity in room. wanted is the opposite of unwanted. person does not want unwanted. person wants light. You would buy because you want to wanted. free is the opposite of buy. free is similar to unoccupied. ')

I believe this can be improved be selecting better paths. I tried to do this by the approach explained below but the it took a lot of time so I think for this I would need a better hardware and GPU. (or at least local so the environment wont restart!)

For finding the path between two given words, I first found the length of the shortest path between the pair of words in the graph. Later, I found all the paths of that length (length of the shortest path between the pair) and choose the best path based on the path score (Sum of the weights). And later as same as the approach previously explained we find the most relevant path among all the pairs.