Choice of Plausible Alternatives

SuperGLUE

Li, Yansong; Qu, Shuzheng; Su, Xuanyu;
Yang, Siyuan; Linkletter, Maurice

uOttawa

SUMMARY

- We chose to do one of the SuperGLUE tasks
- Task: Choice of Plausible Alternatives (COPA)
- Our approach is based on the BERT language model
- BERT is an open source language model created by Google
- We have improved our model's accuracy from 60.4% to 74.7% using an Agile development process.
- By pre-training our model on a large set of MultiNLI data we have been able to improve our model's accuracy by 2%
- For future work we would like to use RoBERTa rather than BERT.

SuperGLUE

https://super.gluebenchmark.com/

COPA

Choice of Plausible Alternatives (COPA)

- The COPA task provides researchers with a tool for assessing progress in open-domain common-sense causal reasoning.
- COPA consists of 1000 questions, split equally into development and test sets of 500 questions each.
- Each question is composed of a premise and two alternatives, where the task is to select the alternative that more plausibly has a causal relation with the premise.
- The correct alternative is randomized so that the expected performance of randomly guessing is 50%.

COPA example:

Premise: The man broke his toe. **Type:** Cause
Alternative 1: He got a hole in his sock.
Alternative 2: He dropped a hammer on his foot.

Premise: I tipped the bottle. **Type:** Effect
Alternative 1: The liquid in the bottle froze.
Alternative 2: The liquid in the bottle poured out.

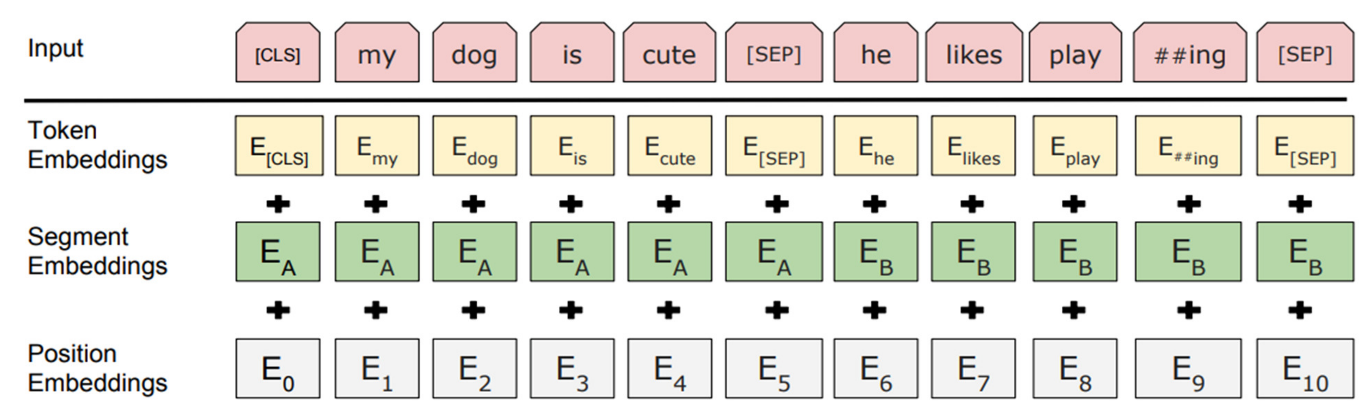
Premise: I knocked on my neighbor's door. **Type:** Effect
Alternative 1: My neighbor invited me in.
Alternative 2: My neighbor left his house.

OUR APPROACH

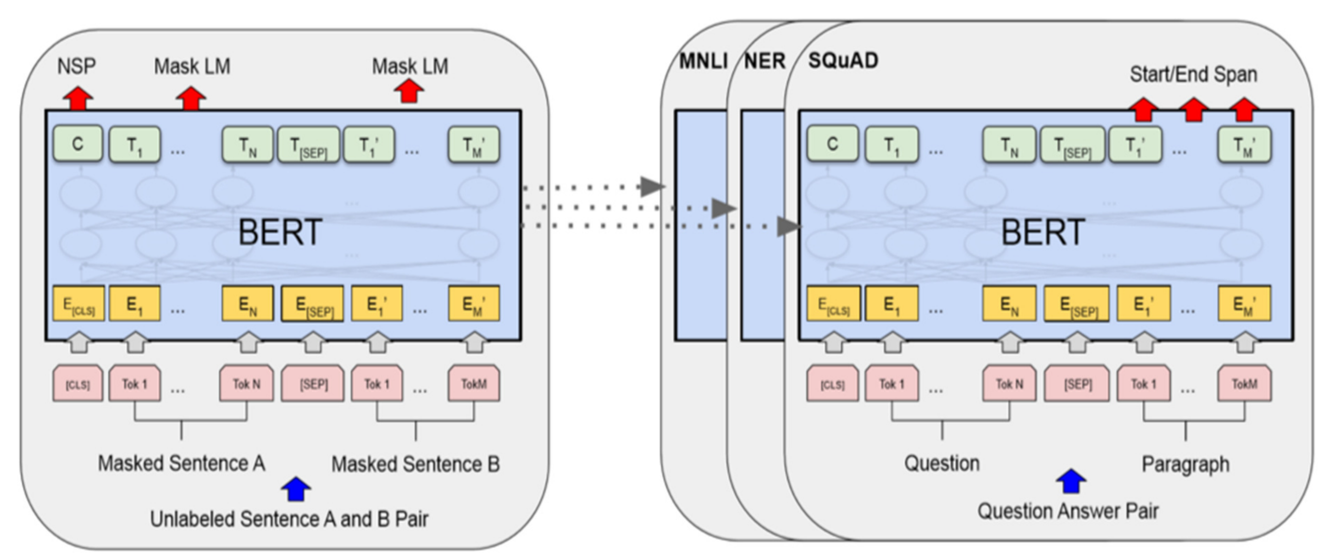
- We researched current state-of-the-art approaches for the COPA task.
- We chose to base our implementation on BERT $_{base}$ pre-trained Language model [3].
- We performed data pre-processing to add additional categorization to our training and evaluation data.
- Our model is comprised of 4 major components.
  - Data pre-processing
  - Linear Transformation layer
  - BERT language model layer(s).
  - A classification layer (learned matrices of weights and biases) with a Softmax output.
- Used an Agile Development process to iterate on our model's implementation.
- Using this Agile process we have been able to increase our model's accuracy from **60.40%** to **74.70%.**
- Generated an additional 10,000 pieces of training data from the MultiNLI dataset [4][2]
- Using the MultiNLI training data we were able to improve our model's accuracy from **72.7%** to **74.7%.**
- We are also attempting to improve our model's accuracy by use WordNet as a in-memory model in data pre-processing. ***This is a work-in-progress and a stretch goal for the project.***

BERT

- BERT (Bidirectional Encoder Representations from Transformers) is an open source language model created by Google.
- Input representation can be composed of three-part Embedding summation [3].

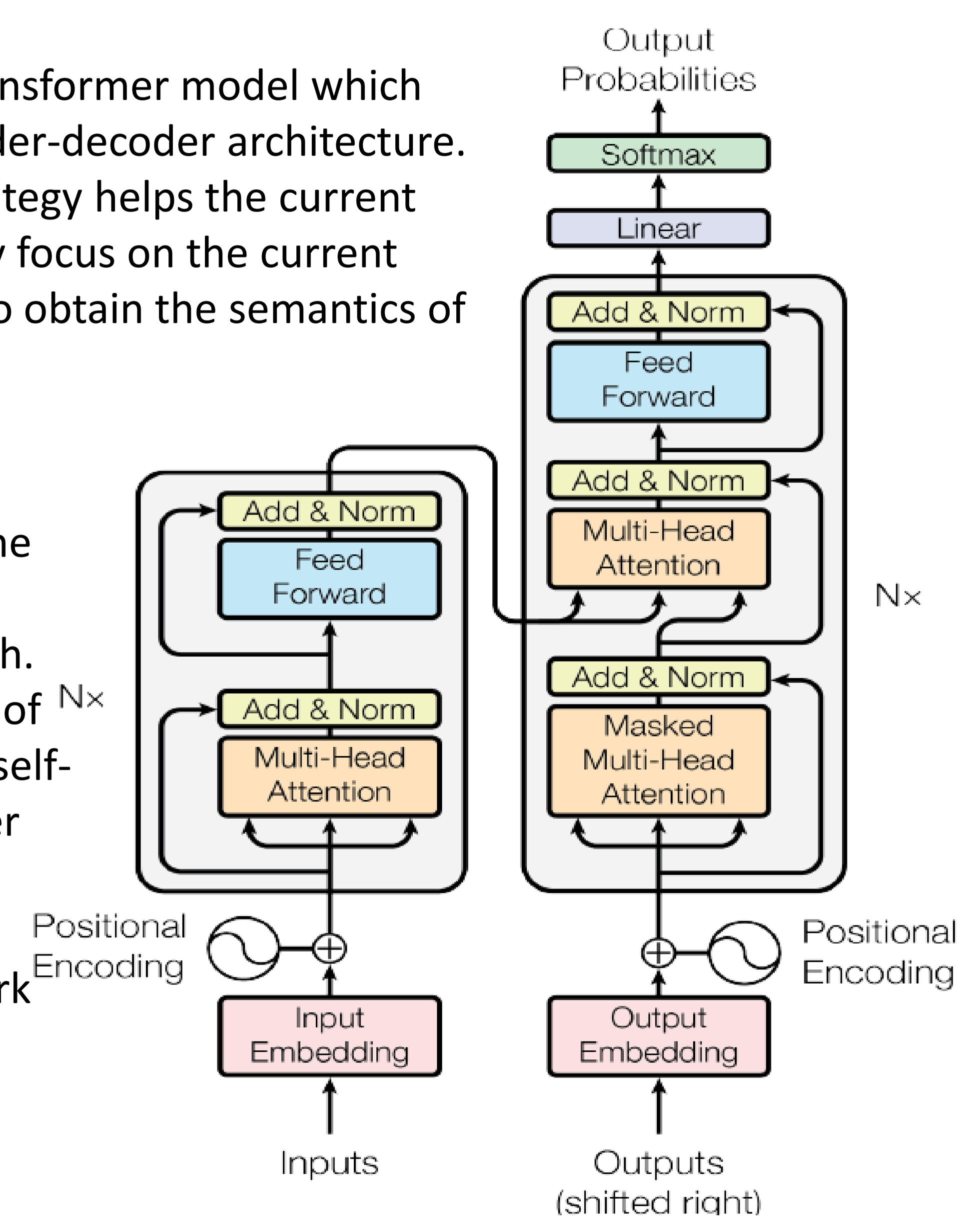


- Pre-training: trained on unlabeled data (unsupervised learning);
- Fine-Tuning: using labeled data from the downstream tasks (COPA).

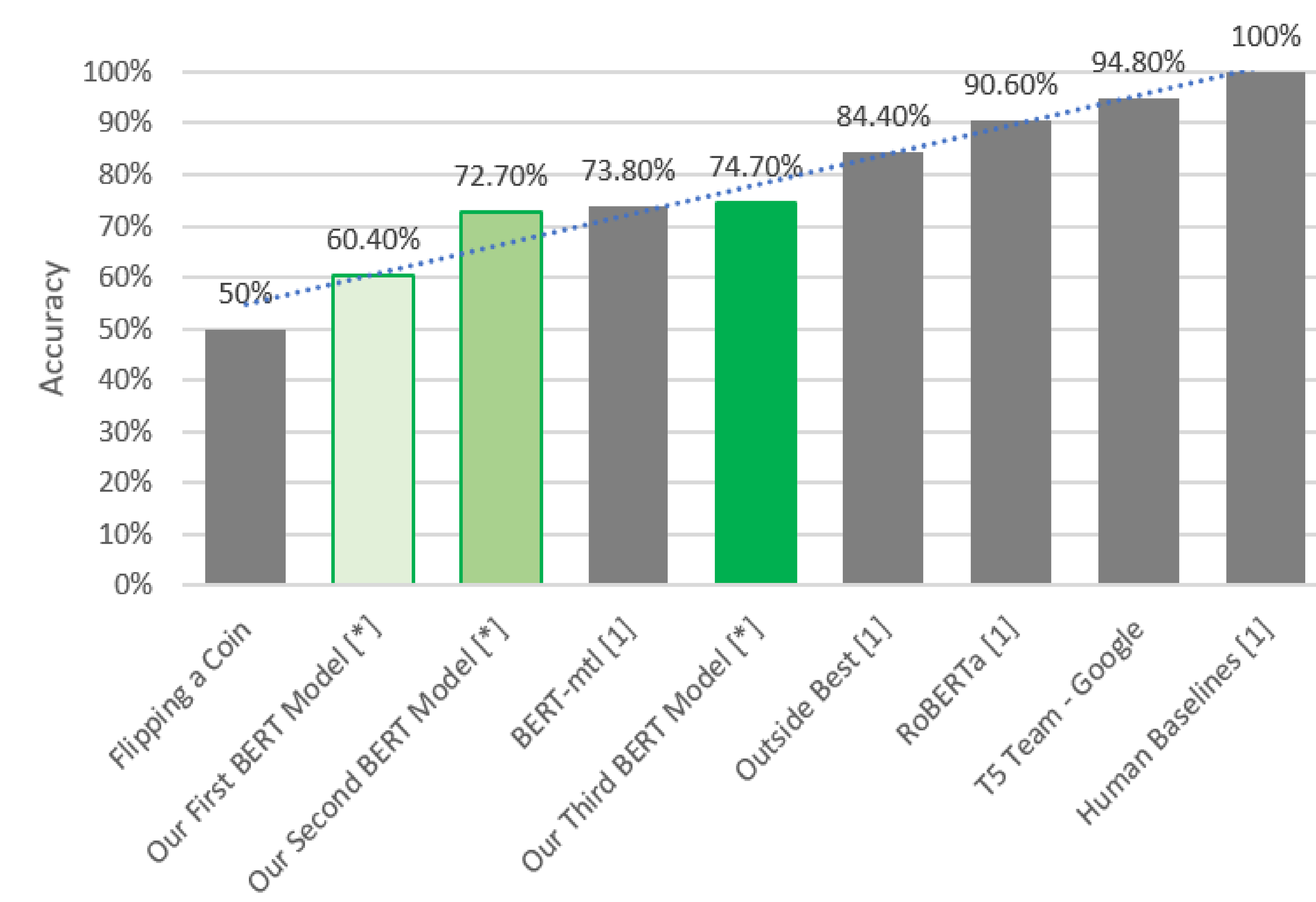


Pre-training Fine-Tuning

- BERT uses transformer model which has the encoder-decoder architecture.
- Attention strategy helps the current node not only focus on the current word, but also obtain the semantics of the context.
- Both encoder and decoder share the same number in network depth.
- Both consists of two layers, a self-attention layer and a feedforward neural network



RESULTS



** Our model's accuracy is based on the average accuracy of 10 consecutive experiments using the same hyper-parameter configuration.*

CONCLUSION

- By pre-training our model with MultiNLI[4][2] data and then fine-tuning on the COPA data our BERT model obtained an accuracy of 74.7% on the COPA task.
- By using an Agile Development process we have been able to rapidly improve our model's accuracy.
- For future work we would like to continue looking at ways to use WordNet, as an in-memory model, in our data pre-processing step.
- Can we uses WordNet's text similarity to enrich our COPA sentence and help our model learn/extract common-sense reasoning?
- For future work we would also consider using the RoBERTa pre-trained language model. RoBERTa is Facebook's open source version of the BERT model. A RoBERTa based model currently holds second place in the COPA task on the SuperGLUE leader board.

REFERENCES

1. Alex Wang, Yada Pruksachatkun, Nikita Nangia, Amanpreet Singh, Julian Michael, Felix Hill, Omer Levy, and Samuel R. Bowman. SuperGLUE: A stickier benchmark for general-purpose language understanding systems. *arXiv preprint arXiv:1905.00537, 2019b*
2. Zhongyang Li, Tongfei Chen, Benjamin Van Durme . Learning to Rank for Plausible Plausibility. *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, pages 4818–4823 Florence, Italy, July 28 - August 2, 2019.*
3. Jacob Devlin and Ming-Wei Chang and Kenton Lee and Kristina Toutanova. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *In Proc. NAACL.*
4. Adina Williams, Nikita Nangia, and Samuel R Bowman. 2017. A broad-coverage challenge corpus for sentence understanding through inference. *arXiv preprint arXiv:1704.05426 .*