# Cross-lingual Transfer learning for Causal Commonsense Reasoning

QIU Jiaxing, XIONG Guangzhi

Dec. 21, 2020

# Background

*Causal Commonsense Reasoning*

- "Bridge between premises and possible hypotheses with world knowledge that is not explicit in text"

| | Premise | Hypothesis |
|---|---|---|
| √ | I poured water on my sleeping friend. | My friend awoke. |
| × | Teddy is my son. | Teddy is a male. |

# Background

*Cross-lingual Transfer Learning*

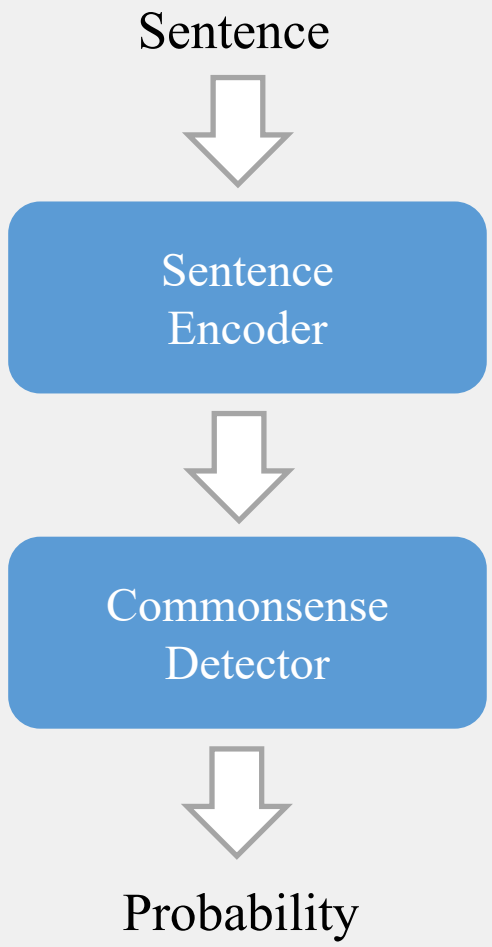
- Solve one problem and apply it to a different but related problem (Transfer Learning)
- Leverage large datasets available in one language—typically English—to build multilingual models that can generalize to other languages
- ZERO-SHOT / FEW-SHOT

# Overview

## PROCEDURE

- Data preprocessing
- Pretraining
- Fine-tuning
- Test

## MODEL



# Datasets

- Datasets for Pretraining
  - **SocialIQA**
    - "context": "Tracy protected her teammates from injury when she saw an accident about to happen prevented it."
    - "question": "Why did Tracy do this?“
    - "answerA": "make a play", "answerB": "prevent injuries", "answerC": "injure them"
  - **WinoGrande**
    - "sentence": "The GPS and map helped me navigate home.  I got lost when the _ got turned off. "
    - "option1": "GPS", "option2": "map"
- Datasets for Fine-tuning
  - **COPA**
    - "premise“: "I poured water on my sleeping friend. "
    - "choice1": "My friend awoke. ", "choice2": "My friend snored. "
- Datasets for Test
  - **COPA**
  - **XCOPA**

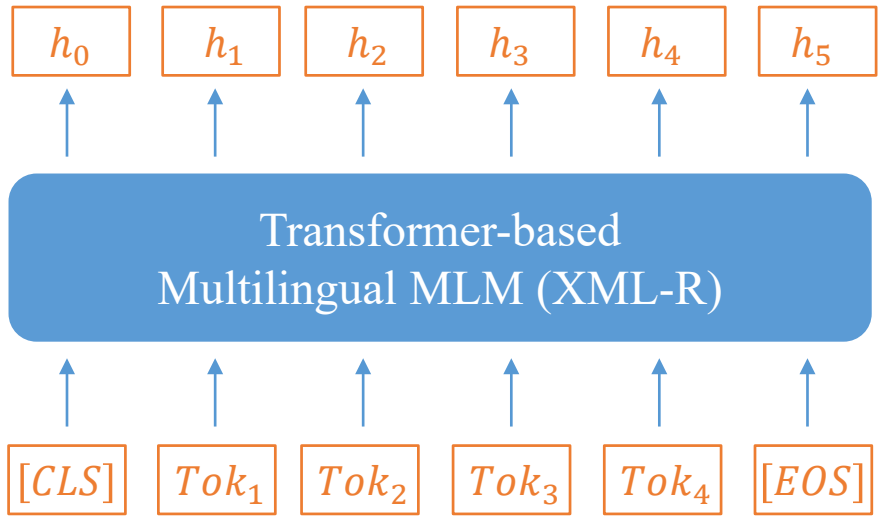
# Methodology

*Sentence Encoder*

- Input sentence
  $S = \{[CLS], Tok_1, Tok_2, \dots, Tok_n, [EOS]\}$

- Hidden states
  $H = \{h_0, h_1, h_2, \dots, h_n, h_{n+1}\}$

- Take the hidden state of $[CLS]$ $(h\_0)$ as the representation of the sentence



# Methodology

## *Commonsense Detector*

- Obtain the representation of the $i$- sentence : $r_i$

- Compute the possibility of having a causal relation

$$P_i = sigmoid(W_b\sigma(W_a r_i + b_a) + b_b)$$

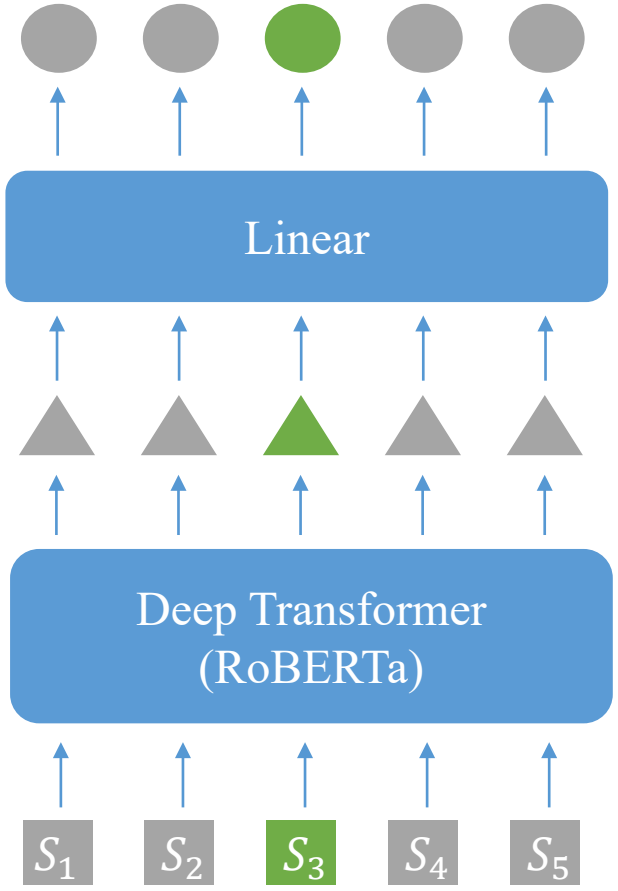
- Use binary cross entropy (BCE) as the detection loss

$$\mathcal{L} = -\frac{1}{N}\sum_{i=1}^{N} \mathbb{I}(label_i = 1)\log P_i + \mathbb{I}(label_i = 0)\log(1 - P_i)$$

# Methodology

*Denoising Module*

- Background
  - Noncausal relation with positive label
- Target
  - Denoise low-quality input data
- Method
  - Train a rank model for monolingual (English) causal commensense reasoning detection
  - Compute positive scores for $N$ sentences
  - Keep top $k$ sentences



# Methodology

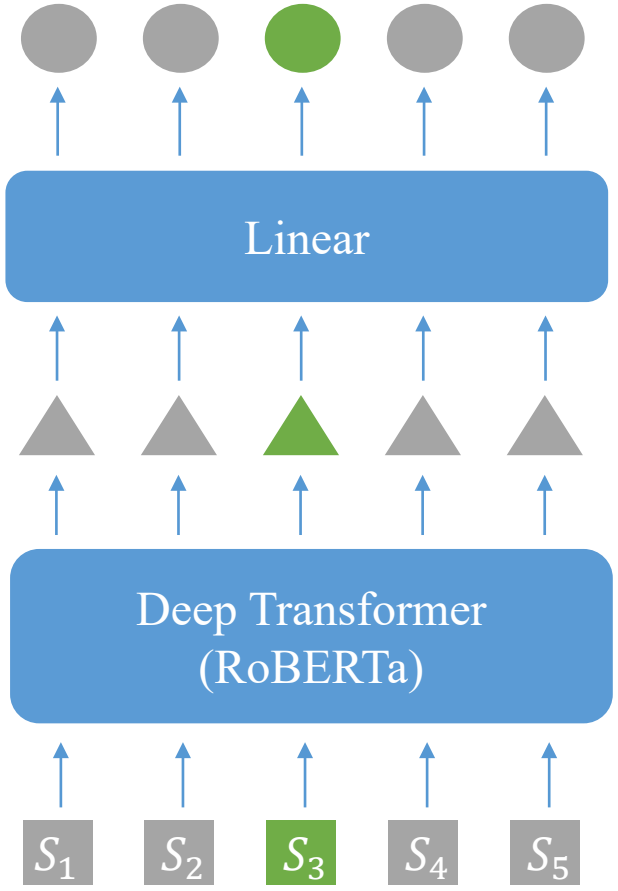
*Denoising Module*

## How to train?

- Input $m$ instances $\{S_i\}_{i=1}^{m}$ where only one is positive
- Compute their positive scores using Roberta and MLP

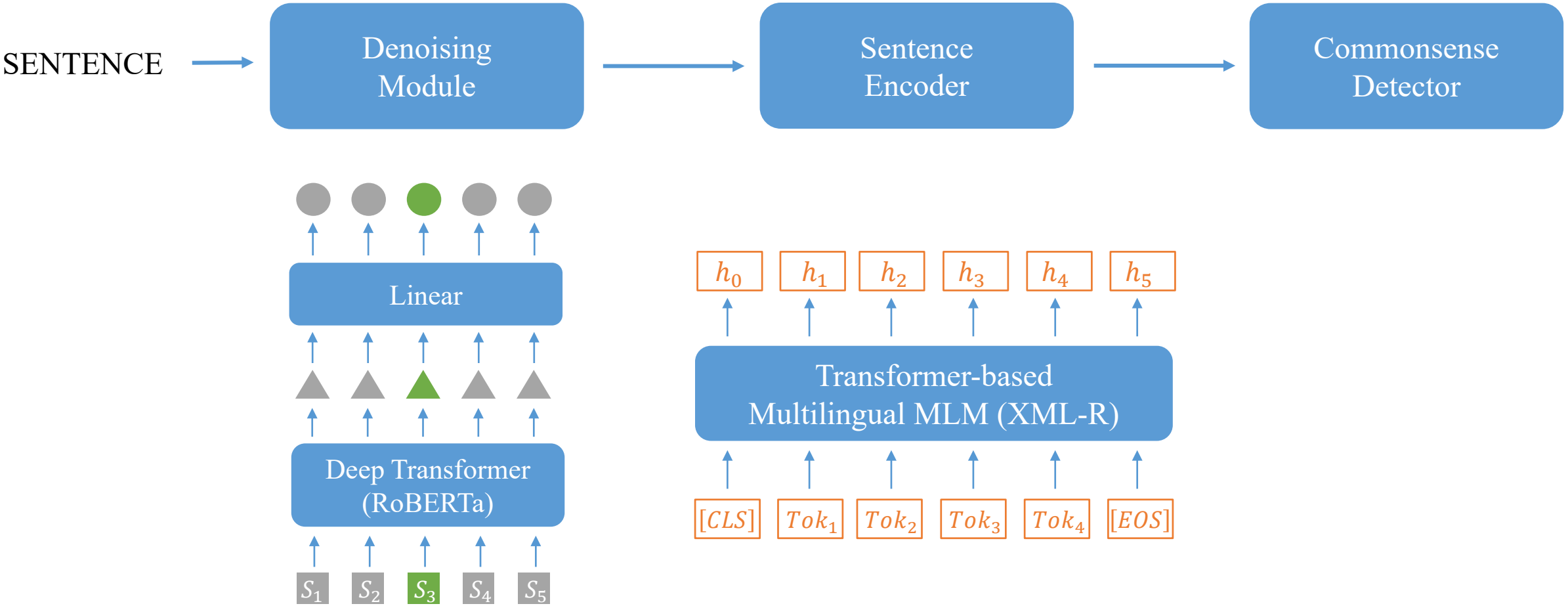
$$r_i = Encoder(S_i)$$
$$score_i = w_m r_i + b_m$$

- Apply a softmax function to compute the probability of $i$-th instance to be positive



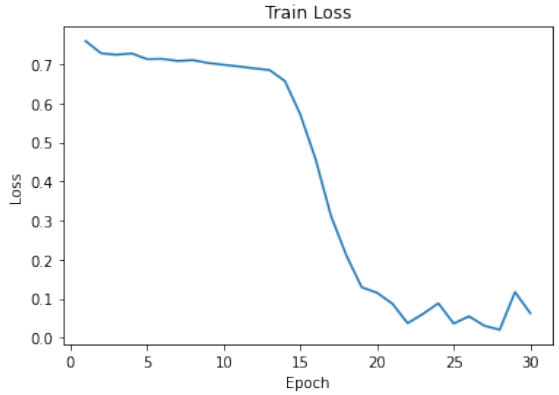
# Methodology

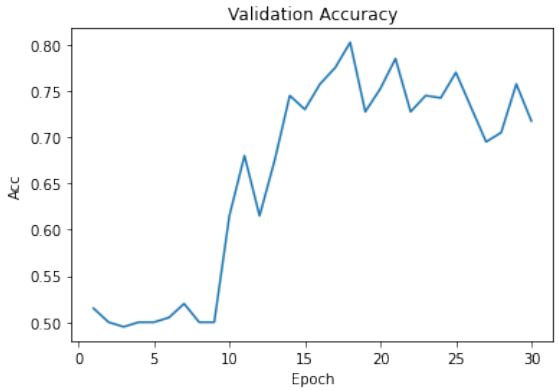
*Network Architecture*

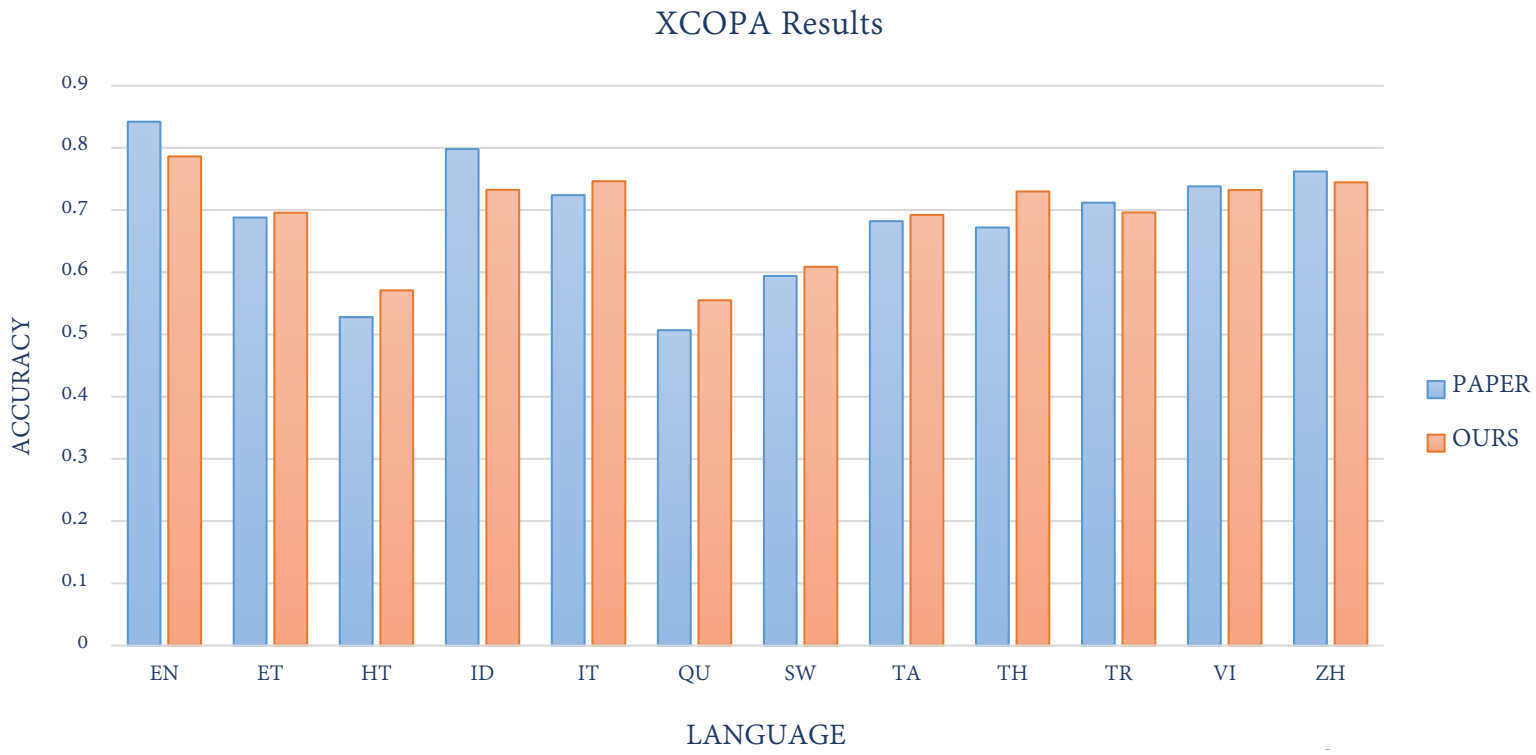


# Experiment

- Pretraining: batch_size=64 Adam with learning_rate=4e-4 and weight_decay=0.01
- Fine tuning: batch_size=16 Adam with learning_rate=1e-5 and weight_decay=0.1



# References


Murray Singer, Michael Halldorson, Jeffrey C Lear, and Peter Andrusiak. 1992. Validation of causal bridging inferences in discourse understanding. Journal of Memory and Language, 31(4):507–524.

Artetxe, M., Labaka, G. and Agirre, E., 2020. Translation Artifacts in Cross-lingual Transfer Learning. arXiv preprint arXiv:2004.04721.

Ponti, E.M., Glavaš, G., Majewska, O., Liu, Q., Vulić, I. and Korhonen, A., 2020. XCOPA: A Multilingual Dataset for Causal Commonsense Reasoning. *arXiv preprint arXiv:2005.00333*.

Conneau, A., Khandelwal, K., Goyal, N., Chaudhary, V., Wenzek, G., Guzmán, F., Grave, E., Ott, M., Zettlemoyer, L. and Stoyanov, V., 2019. Unsupervised cross-lingual representation learning at scale. *arXiv preprint arXiv:1911.02116*.

Roemmele, M., Bejan, C.A. and Gordon, A.S., 2011, March. Choice of Plausible Alternatives: An Evaluation of Commonsense Causal Reasoning. In AAAI spring symposium: logical formalizations of commonsense reasoning (pp. 90-95).

Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L. and Stoyanov, V., 2019. Roberta: A robustly optimized bert pretraining approach. arXiv preprint arXiv:1907.11692

Sap, M., Rashkin, H., Chen, D., LeBras, R. and Choi, Y., 2019. Socialiqa: Commonsense reasoning about social interactions. arXiv preprint arXiv:1904.09728.

Sakaguchi, K., Le Bras, R., Bhagavatula, C. and Choi, Y., 2020, April. Winogrande: An adversarial winograd schema challenge at scale. In *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 34, No. 05, pp. 8732-8740).

# Thanks for your attention