Multilingual PIQA Dataset Submission

Anonymous ACL submission

Abstract

Multilingual reasoning is important. We contribute towards the data collection effort of the MRL 2025 shared task by collecting 125 instances of question answering pairs in 5 languages in the style of PIQA (Bisk et al., 2020). All examples are manually constructed and validated by at another native speaker of the language. We hope our effort is a part of a small step towards building AI systems that can reason in diverse cultural and linguistic settings.

1 Method

We aim to incorporate as much culturally-relevant customs, objects, and locations as possible in the datapoints we generate. With this goal in mind, we first think of a culturally-specific item. We then brainstorm about physical properties of such item that can be incorporated into a PIQA-style datapoint. For example, chopsticks are traditional utensils used in China and are usually made of wood. We therefore can ask models to predict whether dropping the chopstick on the ground will resulting in chopsticks breaking.

2 Dataset Statistics

We manually come up with a total of 125 datapoints in Bengali, Chinese, Korean, and Turkish. Each data-point is manually checked by another native speaker of that language. For Chinese, we have 20 questions in Taiwanese Chinese (in traditional Han characters), the rest are mainland Chinese (in simplified characters).

3 Conclusions

We submit our contribution to the shared task. We hope our contributions are useful for the overall goal of improving physical reasoning capabilities of modern language model in diverse languages.

Language	Count
Bengali	15
Chinese	69
Greek	21
Korean	11
Turkish	9
Total	125

Table 1: Counts of datapoints for each language

References

Yonatan Bisk, Rowan Zellers, Jianfeng Gao, Yejin Choi, and 1 others. 2020. Piqa: Reasoning about physical commonsense in natural language. In *Proceedings of the AAAI conference on artificial intelligence*, volume 34, pages 7432–7439.